THE IMPACT OF IGNORING A LEVEL OF NESTING STRUCTURE IN MULTILEVEL GROWTH MIXTURE MODEL: A MONTE CARLO STUDY

A Dissertation

by

QI CHEN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2008

Major Subject: Educational Psychology

THE IMPACT OF IGNORING A LEVEL OF NESTING STRUCTURE IN MULTILEVEL GROWTH MIXTURE MODEL: A MONTE CARLO STUDY

A Dissertation

by

QI CHEN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Co-Chairs of Committee,	Oi-Man Kwok
	Victor L. Willson
Committee Members,	Robert J. Hall
	F. Michael Speed
Head of Department,	Michael Benz

August 2008

Major Subject: Educational Psychology

ABSTRACT

The Impact of Ignoring a Level of Nesting Structure in Multilevel Growth Mixture Model: A Monte Carlo Study. (August 2008)

Qi Chen, B.A., Nanjing University; M.A., Nanjing University

Co-Chairs of Advisory Committee: Dr. Oi-Man Kwok
Dr. Victor L. Willson

The number of longitudinal studies has increased steadily in various social science disciplines over the last decade. Growth Mixture Modeling (GMM) has emerged among the new approaches for analyzing longitudinal data. It can be viewed as a combination of Hierarchical Linear Modeling, Latent Growth Curve Modeling and Finite Mixture Modeling. The combination of both continuous and categorical latent variables makes GMM a flexible analysis procedure. However, when researchers analyze their data using GMM, some may assume that the units are independent of each other even though it may not always be the case. The purpose of this dissertation was to examine the impact of ignoring a higher nesting structure in Multilevel Growth Mixture Modeling on the accuracy of classification of individuals and the accuracy on tests of significance (i.e., Type I error rate and statistical power) of the parameter estimates for the model in each subpopulation. Two simulation studies were conducted. In the first study, the impact of misspecifying the multilevel mixture model is investigated by ignoring a level of nesting structure in cross-sectional data. In the second study, longitudinal clustered data (e.g., repeated measures nested within units and units nested within clusters) are analyzed correctly and with a misspecification ignoring the highest level of the nesting structure.

Results indicate that ignoring a higher level nesting structure results in lower classification accuracy, less accurate fixed effect estimates, inflation of lower-level variance estimates, and less accurate standard error estimates, the latter result which in turn affects the accuracy of tests of significance for the fixed effects. The magnitude of the intra-class correlation (ICC) coefficient has a substantial impact when a higher level nesting structure is ignored; the higher the ICC, the more variance at the highest level is ignored, and the worse the performance of the model. The implication for applied researchers is that it is important to model the multilevel data structure in (growth) mixture modeling. In addition, researchers should be cautious in interpreting their results if ignoring a higher level nesting structure is inevitable. Limitations concerning appropriate use of latent class analysis in growth modeling include unknown effects of incorrect estimation of the number of latent classes, non-normal distribution effects, and different growth patterns within-group and between-group.

DEDICATION

To my mother, Zuying Chen, who loves me unconditionally, devotes everything she has to me, and has gone through many difficulties in life but has never compromised on the education I received.

ACKNOWLEDGEMENTS

I first would like to acknowledge the members of my guidance committee. I want to express my deep gratitude to Dr. Oi-Man Kwok and Dr. Victor Willson for their supervision during the past three years. It is with their patient guidance, incisive comments, and full support that I came to know my dissertation topic and finished the challenging research. I learned tremendously from both of them on how to do research. More importantly, they are exemplary role models for my academic journey. I also want to specially thank Dr. Robert Hall for his continuing encouragement and help when I was facing difficulties in my life and research. Finally, I want to thank Dr. Michael Speed for his indispensable advice for this dissertation.

I would like to express my heartfelt thanks to Dr. Jan Hughes. The empirical research work with Dr. Hughes familiarized me with the world of child development and opened my horizon of quantitative methods. Most importantly, Dr. Hughes had been a mentor with love and warmth for me in both career development and my personal life.

Thanks also go to my friends and colleagues and the department faculty and staff for making my time at Aggieland a memorable experience. I want to thank Dr. Karen Murphy for being my mentor and "American mom" at my puzzled moments in life, Dr. Wen Luo for sharing her inspirational thoughts and providing help during the course of this research, and Dr. Duan Zhang for sharing her experiences and encouragement when I struggled. I especially want to thank all my professors for being accommodating to my family situation, so that I could work out a way to balance my study and family life. I cannot imagine completing this dissertation without their care and support.

A special word of thank is due to my friends at the International Christian Fellowship. Their love helped me overcome the loneliness and hang in there whenever I wanted to give up during the difficult times when my husband was studying in Los Angeles. I also want to extend my thanks to my longtime teacher and friend Ms. Judy Pehrson for always encouraging me to pursue graduate studies in the US, my longtime friends Jing Bai and Fang Deng, for sharing our dreams since we were in high school and encouraging me all the time.

Finally, I want to thank my mother Zuying Chen, my father Shoujin Chen, my sister Jiangrong Chen and all my relatives, for their unconditional love and support in the past thirty years. I also want to thank my uncle Ying Chen in Taiwan, who always encourages me to get the highest level of education and supports me. Last, but not the least, I want to thank my dear husband Xiaochen Xu for his hard work for the family, for always being my harbor when I got tired or discouraged, his understanding, trust, support, and love.

TABLE OF CONTENTS

		Page
ABSTRACT		iii
DEDICATION		v
ACKNOWLEDGEMENTS		vi
TABLE OF CONTENTS		viii
LIST OF TABLES		x
LIST OF FIGURES		xi
CHAPTER		
I	INTRODUCTION AND LITERATURE REVIEW	1
	Importance of GMM in Developmental Research	1
	Difference of GMM from Other Models	3
	Concepts Related to Growth Mixture Models	7
	The Necessity of Studying MGMM	31
II	STUDY ONE	34
	Simulation 1A	34
	Simulation 1B	42
	Discussion	46
III	STUDY TWO	50
	Method	50
	Results	56
	Discussion	61
IV	GENERAL DISCUSSION AND CONCLUSIONS	65
	Summary and Discussion	65
	Implications	68
	Limitations and Suggestions for Future Research	69

Page

REFERENCES 70

APPENDIX A TABLES 79

APPENDIX B FIGURES 95

VITA 111

LIST OF TABLES

TABLE		Page
1	Hit Rate of True and False Models in Simulation 1A	80
2	Relative Bias of Group Mean Estimates in Simulation 1A	81
3	Relative Bias of Variance Estimates in Simulation 1A	82
4	Relative Bias of Standard Errors of Group Mean Estimates in Simulation 1A	83
5	Hit Rate of True and False Models in Simulation 1B	84
6	Relative Bias of Group Mean Estimates in Simulation 1B	85
7	Relative Bias of Variance Estimates in Simulation 1B	86
8	Relative Bias of Standard Errors of Group Mean Estimates in Simulation 1B	87
9	Mean Hit Rate for True and False Models	88
10	Mean Relative Bias of Fixed Effect Estimates for True and False Models	89
11	Mean Relative Bias of Level-1 Residual Variance Estimates for True and False Models	90
12	Mean Relative Bias of Level-2 Variance and Covariance Estimates for True and False Models	91
13	Mean Relative Bias of Standard Errors of Fixed Effect Estimates for False Models	92
14	Mean Power to Detect Significant Level-2 Variance and Covariance for True and False Models	93
15	$\eta^2 \geq .01$ Values from 5-Way ANOVA for True and False Models	94

LIST OF FIGURES

FIGURE		Page
1	Diagram Illustrating Relationships between Models	96
2	Linear Latent Growth Curve Model with Four Repeated Measures and One Covariate	97
3	Linear Growth Mixture Model with Four Repeated Measures and One Covariate Predicting Both Growth Factors and Latent Class	98
4	Interaction Effect between ICC and R^2 for Estimated Mean Hit Rate Difference between True and False Models	99
5	Interaction Effect between ICC and R^2 for Estimated Mean Relative Bias for Level-1 Variance Estimate under False Model	100
6	Interaction Effect between Imbalance Type and R^2 for Estimated Mean Hit Rate Difference under True and False Models	101
7	Interaction Effect between Imbalance Type and R^2 for Estimated Mean Relative Bias for Class 1 under True Model	102
8	Interaction Effect between Imbalance Type and R^2 for Estimated Mean Relative Bias for Class 2 under True Model	103
9	Interaction Effect between Imbalance Type and R^2 for Estimated Mean Relative Bias for Class 1 under False Model	104
10	Interaction Effect between Imbalance Type and R^2 for Estimated Mean Relative Bias for Class 2 under False Model	105
11	Interaction Effect between Imbalance Type and R^2 for Estimated Mean Relative Bias for Class 1 Level-1 Variance under True Model	106
12	Interaction Effect between Imbalance Type and R^2 for Estimated Mean Relative Bias for Class 2 Level-1 Variance under True Model	107
13	Interaction Effect between Imbalance Type and R^2 for Estimated Mean Relative Bias for Class 1 Level-1 Variance under False Model	108

FIGURE Page

14 Interaction Effect between Imbalance Type and R^2 for Estimated Mean Relative Bias for Class 2 Level-1 Variance under False Model 109

15 Interaction Effect between ICC and τ_π for Estimated Mean Relative Bias of Intercept Variance under False Model 110

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

The following chapter is composed of four major parts related to growth mixture models (GMM). The first part reviews the emergence and the importance of GMM in developmental research, the second part distinguishes a number of concepts that is related to GMM, the third part introduces the basic concepts, models, and limitations related to GMM (i.e., Finite Mixture Models, Latent Growth Curve Models, Growth Mixture Models, and Multilevel Growth Mixture Models), and the last part reviews research related to multilevel model and the necessity of studying MGMM.

Importance of GMM in Developmental Research

The number of longitudinal studies increases steadily in different social science disciplines over the last decade (Khoo, West, Wu, & Kwok, 2006; West, Biesanz, & Kwok, 2003). Many new approaches have been proposed to analyze longitudinal data, including hierarchical linear models (HLM, Raudenbush & Bryk, 2002), structural equation based latent growth curve modeling (LGCM, Meredith & Tisak, 1990; Bollen & Curran, 2006; Duncan, Duncan, Strycker, Li, & Alpert, 1999), latent class growth analysis (LCGA, Nagin, 1999), and growth mixture modeling (GMM, Muthén, 2004).

Among these approaches, both HLM and LGCM are variable-centered approaches to data analysis (Muthén& Muthén, 2000; Connell & Frye, 2006) because the goals of these models are to examine the relations between independent and dependent variables and study how constructs influence their indicators. In contrast to the variable-centered approach, the person-centered approach aims at classifying individuals into different

---

This dissertation follows the style of *Journal of Educational Psychology*.

categories. That is, individuals who shared similar characteristics are classified into the same category/group (Muthén & Muthén, 2000). GMM and LCGA are recent developments in statistical technique to handle the longitudinal data integrating both variable- and person-centered approaches. These approaches can identify the unobserved and heterogeneous developmental trajectories, which are represented by the latent categorical/grouping variable (Muthén & Muthén, 2000). The relationship between growth parameters can also be examined within each latent class of individuals. The relationship between GMM, LGCM, and LCGA will be discussed in detail later in the chapter.

Growth Mixture Models (GMM) has been recently receiving more attention in educational and psychological research. One major reason of this development is due to the flexibility of GMM. For example, GMM allows more complex ideas of development to be examined because it can be a part of the general latent variable model including other factors (i.e. factor for covariates with measurement errors) influencing the growth factors or outcomes influenced by the growth process. Moreover, as pointed out by Nagin (1999), psychology has a long tradition of theorizing development based on groups in different research areas, such as theories of different kinds of personality development, and theories about prosocial and antisocial behaviors development. These theories raise questions such as whether one type of people tends to have certain developmental trajectories distinguishable from other types of people. It is not likely that population differences in developmental trajectories will be as clearly distinct as the differences found in biological or physical phenomena (i.e. different types of animal species or the four states of matter). Nevertheless, GMM, as a statistical method, can be used to test theories that predict prototypal developmental trajectories in population. Rather than assigning subjective

categorization, GMM provides a more objective method for determining the number of latent classes and calculating the posterior probability of class membership for each individual.

Difference of GMM from Other Models

There are many person-centered statistical approaches addressing the unobserved heterogeneity within data, such as cluster analysis, finite mixture models, latent class analysis, latent transition analysis, and GMM. The purpose of this section is to clarify the similarity and difference between these approaches, especially to distinguish GMM from all these approaches.

Cluster analysis (Everitt, Landau & Leese, 2001) is a generic term for a wide range of numerical methods for examining multivariate data with a view to uncovering or discovering groups or clusters from the observed data. Data clustering algorithms can be either hierarchical or partitional. Hierarchical clustering is a type of cluster analysis in which data are not partitioned into a particular number of classes or clusters at a single step. Instead, the classification consists of a series of partitions, which may run from a single cluster containing all individuals to *n* clusters each containing a single individual. Hierarchical clustering can use either the agglomerative methods or the divisive methods. The researcher needs to decide when to stop for an optimal number of clusters. This type of cluster analysis is mostly relevant in biological applications, studies of social systems, museology, and librarianship where hierarchy is implicit in the subject matter. On the other hand, when a particular partition is required and there is no underlying hierarchy, partitional algorithms or optimization clustering techniques that determine the number of clusters at one point is more appropriate. Optimization clustering techniques are a class of

clustering techniques which produce partition of the individuals into a specified number of groups by either minimizing or maximizing some numerical criterion.

Finite Mixture Models (FMM), also called Mixture Modeling (Muthén& Muthén, 2006) and Latent Class Cluster Analysis (Vermunt & Magidson, 2002), is one type of partitional cluster analysis (Everitt et al., 2001). More detailed information about FMM will be provided in later sections. Basically, Finite Mixture Models assume that parameters of a statistical model of interest differ across unobserved subgroups called latent classes or mixture components. In FMM, these latent/unobserved classes or subpopulations are represented by the categorical latent variables in the model (Muthén & Muthén, 2006). It aims at finding a model that can minimize the within-class homogeneity and maximize the across-class heterogeneity. As a model-based clustering approach, FMM differs from standard cluster analysis (Everitt et al., 2001). Moreover, FMM can be applied to different kinds of data, including continuous and categorical, cross-sectional and longitudinal data.

There are usually two major parts in FMM: one part is the observed indicators (also called dependent variables, outcome variables, outputs, endogenous variables, and items) measured for each individual, and the other part is the latent categorical variable (also called exogenous variable) that specifies the unobserved group membership for each individual (Vermunt & Magidson, 2002). FMM can be applied to both cross-sectional and longitudinal data. Traditionally, Latent Class Analysis (LCA), also referred as Latent Class Models (LCM) (Clogg, 1995), is used when the observed indicators are categorical and cross-sectional. For instance, the indicators may correspond to a set of dichotomous diagnostic criteria items, and the latent categorical variable may describe the presence or absence of an alcohol use disorder (Muthén & Muthén, 2000). On the other hand, when

FMM is applied to cross-sectional data with continuous observed indicators, it is called Latent Profile Analysis (LPA) (Lazarsfeld & Henry, 1968).

When FMM is applied to longitudinal data, depending on the kind of change being analyzed, there are three major types of models. Khoo et al. (2006) reviewed three major types of longitudinal models, namely, autoregressive model, latent trait-state model, and growth curve model. Courvoisier, Eid, and Nussbeck (2007) pointed out that these three models analyze two types of change, namely, the long lasting and irreversible change (i.e., autoregressive model and growth curve model) and the change with status alteration over time (i.e., latent trait-state model). When mixture models are applied to analyze longitudinal change, there are three types of models, namely, Latent Transition Analysis (LTA), Mixture Latent State-Trait Analysis (MLSTA), and Growth Mixture Models (GMM).

LTA is applied when the observed indicators are categorical and longitudinal and the aim of the analysis is to detect the qualitative or status change between stages of time (Collins & Flaherty, 2002). In LTA, the latent categorical variable becomes a stage-sequential dynamic latent variable, which changes in systematic ways over time so that each stage represents a qualitatively different way of organizing information. The researcher can specify the number of consistently transitional stages in a model and study the probability of transiting from one class at one time point to another class at the next time point. LTA is mostly used in public health research (Collins & Flaherty, 2002). On the other hand, MLSTA can be viewed as the counterpart of LTA because it deals with continuous rather than categorical observed variables (Courvoisier et al., 2007). In addition to modeling change by separating the stable, occasion-specific, and error-specific

influences, MLSTA attempts to identify latent groups of individuals who differ in the degrees of behavior or attitude and thus susceptible to different levels of influences by occasions of measurement.

When FMM is applied to longitudinal data with irreversible change, it is called Growth Mixture Modeling (GMM). GMM can model either all categorical or all continuous observed variables as well as data with mixed-mode (i.e. both categorical and continuous) variables (Muthén, 2004; Muthén & Muthén, 2006). The focus of this dissertation is on GMM with all continuous observed variables. To simplify the presentation, I will use GMM to stand for the growth mixture models with continuous observed variables. More details on GMM will be discussed in the following sections.

Figure 1 shows the graphical relationship between the different types of analyses presented previously. There are three different size circles labeled as circles 1, 2, and 3 to represent the large, medium and small circles respectively. The largest circle represents all kinds of cluster analysis techniques, the 2nd largest circle represents partitioned clustering, and the area between circles 1 and 2 represents hierarchical clustering. Within circle 2 is circle 3, which represents all kinds of finite mixture models. Within circle 3, on the right side of the solid line are methods for cross sectional data, including LCA for categorical observed indicators, LPA for continuous observed indicators, and FMM for mixed mode data. On the left side of the solid line are FMM for longitudinal data, including LTA for categorical observed indicators, MLSTA for continuous observed indicators, and GMM for mixed-mode data. In the following section, important concepts closely related to the focus of this dissertation will be introduced, namely, Finite Mixture Models for continuous

indicators, Growth Mixture Models for continuous indicators, and Multilevel Growth Mixture Models.

Muthén (2001b, 2002, 2008) and Vermunt and Magidson (2004b, 2004c) present a general latent variable framework, which includes many of the previously mentioned models and some other models with continuous latent variable for the purpose of classifying observed outcomes (e.g., factor analysis for classifying continuous outcomes and latent trait analysis for classifying categorical outcomes). The general framework connects many statistical approaches and reflects Muthén's (2002) aim of integrating different types of statistical and psychometric models into a unifying framework for statistical modeling. Indeed, *Mplus*, the statistical modeling program developed by Muthén & Muthén (2006) is gaining more and more attention in various research fields due to its flexibility of modeling both continuous and categorical latent variables for longitudinal data (Muthén, 2004). Since the focus of this dissertation is not on integrating different methods into a general framework, the details about this framework and the many old and new mixture models derived from the incorporation of the traditional variable-centered analyses with FMM (described in details by Muthén, 2001a, 2001b, 2002, 2004, 2006) will not be elaborated here.

Concepts Related to Growth Mixture Models

Growth Mixture Model is a modeling approach integrating two statistical modeling techniques, namely, Finite Mixture Models and Latent Growth Curve Models. Therefore, in the following section, key concepts related to these two different modeling techniques will be presented, followed by the concept of Growth Mixture Models. Notice however, only the concepts closely related to the focus of this dissertation will be introduced,

namely, Finite Normal Mixture Models for continuous indicators and Growth Mixture Models continuous indicators.

Finite Mixture Model

Due to the availability of inexpensive high-speed computers in the late twenty's century and the advances in posterior simulation techniques, mixture model has received more and more attention among both practitioners and statisticians (McLachlan & Peel, 2000). Finite mixture models (FMM) underpins a number of statistical techniques, one of which is growth mixture modeling (GMM), a technique becoming increasingly useful in longitudinal studies (Muthén & Shedden, 1999; Muthén, 2004; Bauer & Curran, 2004).

The following section on the basic definition and formula of FMM is a summary of the works by Everitt and Hand (1981), McLachlan and Basford (1988), McLachlan & Peel (2000), and Vermunt and Magidson (2005). Finite normal mixture models assume that a population consists of a finite number of unobserved or latent component distributions, each of which characterized by its own normal distribution for the continuous measures. The probability density function (*pdf*) of the finite multivariate normal mixture model is

$$f(y_j;\Psi) = \sum_{i=1}^{g} \pi_i f_i(y_j;\mu_i,\Sigma_i) \qquad (1a)$$

where y_j denotes the observed value of individual j (j = 1, …, N) on the p-th continuous random variable (of a total of P variables where P=1,…, p) and therefore y_j is a $N \times P$ data matrix, Ψ is a vector containing all the unknown parameters in the mixture model. Each y_j can be viewed as arising from a superpopulation G, which is a mixture of a finite

number of g populations (i.e., $G_1,...,G_g$) with the corresponding proportions (i.e., $\pi_1,...,\pi_g$). π_i , referred as *mixing proportions* or *weights*, is the probability of belonging to (latent) class *i* (where $\sum_{i=1}^{g}\pi_i = 1$ and $\pi_i \geq 0$, $i = 1, \ldots, G$). The class specific densities $f_i(y_j)$ are called the *component densities* of the mixture, and specified as multivariate normal with mean vector μ_i and covariance matrix Σ_i. Its PDF is as follows:

$$f_i(y_j;\mu_i,\Sigma_i) = (2\pi)^{-\frac{p}{2}}\left|\Sigma_i\right|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left(y_j-\mu_i\right)^T\Sigma_i^{-1}\left(y_j-\mu_i\right)\right\} \quad (1b)$$

The likelihood function for a sample of $j = 1, \ldots, N$ randomly drawn observations from the mixture is :

$$L(\Psi) = \prod_{j=1}^{N} f(y_j;\Psi) \quad (1c)$$

In the most general specification where no restrictions are imposed on μ_i and Σ_i, a multivariate normal mixture model with *G* latent classes contains *G-1* unknown class proportions, $G \times p$ class-specific means, $G \times p$ class-specific variance, and $G \times p \times (p-1)/2$ class-specific covariances.

Maximum likelihood (ML) is one of the main estimation methods in FMM. Most software packages use the EM algorithm or some modifications of it to find the ML estimates. Upon achieving convergence, the estimates of the model parameters and their

asymptotic covariances are provided. In addition, classification of individuals into different (latent) clusters can be done based on the posterior class membership. The posterior probability of y_j belonging to cluster *i* can be computed by using Bayes' rule:

$$\tau_i(y_j;\Psi)=\frac{\pi_i f_i(y_j;\mu_i,\Sigma_i)}{\sum_{h=1}^{g}\pi_h f_h(y_j;\mu_h,\Sigma_h)} \tag{1d}$$

The standard classification method is to assign individual to the class with the highest posterior probability.

By assuming the data under analysis are composed of a discrete number of components, finite mixture models can handle situations where a single parametric family is unable to provide a satisfactory model for local variations in the observed data (McLachlan & Peel, 2000). FMM is the analytical basis for GMM, in which the population being analyzed is composed of individuals coming from a finite number of latent trajectory classes (Muthén, 2001b). Mixture modeling is similar to multiple group analysis. However, an important difference between mixture modeling and the standard multiple group analysis is that in mixture modeling, the group membership cannot be directly observed (Muthén, 2001b; Vermunt & Magidson, 2005). This is why sometimes researchers refer FMM as latent class analysis (LCA). As mentioned by Muthén (2002) and Vermunt (2007), mainstream statistics often refer mixture models with continuous indicators as finite mixture models and reserve the term LCA for mixture models with all categorical response variables. In this dissertation I will use the two terms interchangeably.

Latent Growth Curve Models

Latent Growth Curve Models (LGCM) is a special form of Structural Equation Modeling (SEM; Bollen & Curran, 2006) to address the analysis of change. The basic idea is that there exist some unobserved latent factors which can account for the means and covariances of (the changes of) the observed repeated measures. The interest of analysis is more in the latent factors, which are referred to as growth factors, instead of the repeated measures.

Bollen and Curran (2006) give a very detailed and organized description of different kinds of LGCM. In their framework, there are unconditional versus conditional LCGM. Conditional LCGM includes covariates affecting the growth trajectory whereas the unconditional LCGM does not. There are linear versus nonlinear LGCM. The linear LGCM is governed by an intercept and a linear slope component whereas the nonlinear LGCM includes higher-order polynomials (i.e., powers of time) that result in quadratic or cubic growth trajectories. There are also univariate versus multivariate LGCM, with the univariate LGCM only considering repeated measures of a specific outcome whereas the multivariate LGCM considering repeated measures for multiple outcomes. In addition, LGCM can be extended to include both continuous and categorical observed outcomes, and in many other ways. Indeed, LGCM is a very useful statistical technique to model longitudinal data. In this dissertation, however, only the notations of the conditional linear LGCM is summarized, on the one hand because the focus of this dissertation is on the linear growth models, on the other hand because this type of models lays the foundation for the other more complicated ones.

First of all, the data requirement for identifying and estimating a polynomial growth trajectory with *d* degrees is to have at least $d+2$ repeated observations (Bollen & Curran, 2006). Therefore, in order to fit a linear trajectory ($d=1$), at least three repeated measures are needed whereas for a quadratic trajectory ($d=2$), at least four measure are needed. Figure 2 shows a hypothetical linear LGCM for four repeated measures and one covariate predicting the intercept and slope.

Unlike traditional SEM model with one factor loading fixed as the scaling indicator for the corresponding factor, LGCM fixes all the loadings to specific values to represent the time intervals. In this figure, α represents the intercept factor with the loadings from the factor to each measure fixed to 1, and β represents the slope factor with the loadings fixed as 0, 1, 2, and 3 for measures from time1 to time4 respectively. The intercepts of the repeated measures are set to 0, and the means for the growth factors are estimated. As a result, the model-implied mean structure of the repeated measures is determined entirely by the means of the growth factors. The estimated residual variance of the repeated measures is the variance not explained by the growth process. In addition to mean, the variances and the covariance of the two growth factors will also be estimated, which reflect the degree of variability of the individual intercepts and slopes around the means.

The conditional linear latent growth curve model would look like this:

Measurement: $$Y_{it} = \alpha_i + \lambda_t \beta_i + \varepsilon_{it} \qquad \varepsilon_{it} \sim N(0, \sigma^2) \qquad (2a)$$

Structural: $$\alpha_i = \mu_\alpha + \sum_{q=1}^{Q} \gamma_{\alpha q} X_{qi} + \zeta_{\alpha i} \qquad \zeta_i \sim N(0, T_\zeta) \qquad (2b)$$

$$\beta_i = \mu_\beta + \sum_{q=1}^{Q} \gamma_{\beta q} X_{qi} + \zeta_{\beta i} \quad (2c)$$

where the level-1 model (i.e., equation 2a) captures the change within individuals over time. Y_{it} is the value of the trajectory variable y for the *i*th individual at time *t* ($t = 1,2,3,...,T$). α_i is the intercept and β_i is the slope for individual *i*. λ_t ($\lambda_t = t - 1$) is a predictor variable measuring time within each individual, and ε_{it} is the residual. The level-2 model (i.e., equations 2b & 2c) captures: A) the overall model with the average initial score (i.e., μ_α) at time 1 ($\lambda_1 = 0$) and the average rate of change over time (i.e., μ_β) across all individuals when the *q* ($q = 1,2,3,...,Q$) time-invariant covariates X_{qi} are zero, B) the variance (or variation) of the initial status (i.e., $V(\zeta_{\alpha i}) = \tau_{00}$) and the variance (or variation) of the rate of change over time (i.e., $V(\zeta_{\alpha i}) = \tau_{11}$) between individuals from the overall model when the covariates X_{qi} are zero, and C) the influence of *q* covariates (reflected by the covariate coefficients *γ*) on the random intercepts and slopes (i.e., the intercepts and slopes will change by $\gamma_{\beta q}$ when there is one unit change in X_q and all other variables are held constant). These models can also be presented in the matrix terms as the following:

Measurement: $$Y_i = \Lambda \eta_i + \varepsilon_i \quad (2d)$$

$$\begin{bmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{iT} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & T-1 \end{bmatrix} \begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} + \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{iT} \end{bmatrix} \tag{2e}$$

$$\mathrm{Y_i} \quad = \quad \Lambda_\mathrm{i} \qquad \eta_\mathrm{i} \quad + \quad \varepsilon_\mathrm{i}$$

Structural:

$$\eta_i = \mu_\eta + \Gamma X_i + \zeta_i \tag{2f}$$

$$\begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} = \begin{bmatrix} \mu_\alpha \\ \mu_\beta \end{bmatrix} + \begin{bmatrix} \gamma_{\alpha 1} & \gamma_{\alpha 2} & \cdots & \gamma_{\alpha Q} \\ \gamma_{\beta 1} & \gamma_{\beta 2} & \cdots & \gamma_{\beta Q} \end{bmatrix} \begin{bmatrix} X_{1i} \\ X_{2i} \\ \vdots \\ X_{Qi} \end{bmatrix} + \begin{bmatrix} \zeta_{\alpha i} \\ \zeta_{\beta i} \end{bmatrix} \tag{2g}$$

$$\eta_i \quad = \quad \mu_\eta \quad + \qquad\qquad \Gamma \qquad\qquad X_i \quad + \quad \zeta_i$$

where for the *i*-th individual, $\mathrm{Y_i}$ contains the observed repeated measures, Λ_i contains the time information, η_i contains the latent intercept and linear growth rate, and ε_i contains the random errors. The means of the latent factors μ_α and μ_β are the average initial status (i.e., $E(\alpha_i) = \mu_\alpha$) and growth rate (i.e., $E(\beta_i) = \mu_\beta$) when covariates are zero, whereas the two ζ_i's are the random effects and the variances of them capture the variances of the initial status (i.e., $V(\zeta_{\alpha i}) = \tau_{00}$) and the rate of change over time (i.e., $V(\zeta_{\alpha i}) = \tau_{11}$).

Γ contains the corresponding regression coefficients of the *Q* exogenous *X*'s predicting the growth factors. Combining equations 2d and 2f, the reduced-form model for LGCM is:

$$Y_i = \Lambda(\mu_\eta + \Gamma X_i) + \Lambda\zeta_i + \varepsilon_i \tag{2h}$$

The above equation allows any number of repeated measures and covariates. In an unconditional linear model, the part related to covariates will be omitted and the model is:

$$Y_i = \Lambda(\mu_\eta + \zeta_i) + \varepsilon_i \tag{2i}$$

LGCM is generally estimated using Maximum Likelihood (ML) estimator and the ML function for each individual in the sample is (Arbuckle, 1996):

$$\ln L_i(\theta) = K_i - \frac{1}{2}\ln\left|\sum_i(\theta)\right| - \frac{1}{2}[z_i - \mu(\theta)]'\sum_i(\theta)[z_i - \mu(\theta)] \tag{2j}$$

where z_i is a vector of observed variables for the *i*th individual, and K_i is a constant unrelated to θ. The likelihood function for all individuals is:

$$\ln L(\theta) = \sum_{i=1}^{N} \ln L_i(\theta) \tag{2k}$$

The ML estimator is quite stable when the observed variables have the same multivariate kurtosis as a multivariate normal distribution (i.e. no excess multivariate kurtosis) (Browne, 1984). Other estimators such as Weighted Least Squares are also available when the multivariate normal distribution assumption is violated. Fit statistics such as CFI, RMSEA, and SRMR are available to assess the fit of the model (Muthén & Muthén, 2006).

It is worth noticing that in Hierarchical Linear Models (HLM) or Multilevel Models (Raudenbush & Bryk, 2002) are very similar models which can also use for

estimating the growth trajectories. A single-level model as expressed in equations 2h under the LGCM framework can be expressed as a two-level model under the HLM framework. The measurement part of LGCM where the growth factors (i.e., α_i & β_i) are measured by the multiple time measures (i.e., Y_{i1} to Y_{it}) can be seen as the level 1 model in HLM, whereas the structural part of LGCM which relates the growth factors to other potential level 2 variables (i.e., time invariant covariates such as gender and SES) becomes the level 2 model in HLM. Although HLM and LGCM are two different approaches for the same problem, LGCM combines the strength of conventional HLM and SEM and has many modeling flexibilities, such as its adaptability in handling measurement error, availability of alternative estimators and fit indices, ability to analyzing the relationship between growth factors, multiple processes and multiple groups, and most important of all, the possibility of being a part of the general latent variable model including other factors (i.e. factor for covariates with measurement errors) influencing the growth factors or outcomes influenced by the growth process (Muthén, 2004; Bollen & Curran, 2006).

Growth Mixture Models

Concept and Notations

Both HLM and LGCM assume that all individuals are drawn from a single population with a common set of population parameters, thus mapping one average trajectory over the individual trajectories. However, in substantive research areas, unobserved/underlying heterogeneous groups often exist, such as in market research (Jedidi, Jagpal & DeSarbo, 1997), developmental psychology (Dolan & van der Maas, 1998), sociology research (Arminger & Stein, 1997), and educational research (Muthén, 2002). Statistical method that address *a priori (or known)* grouping such as multiple group

SEM analysis is not suitable for such situations, because the group is not observed and need to be decided based on the observed measures (Jedidi et al., 1997). Mixture SEM model is thus developed to address the statistical modeling difficulties, such as using a model-based approach to perform cluster analysis concurrently with taking into account measurement error and testing the hypothesized model structure (Jedidi et al., 1997). Mixture SEM is viewed as generalizing or extending the multiple group SEM analysis to where the group membership is unknown (Jedidi et al., 1997; Vermunt & Magidson, 2005). Growth Mixture Models, therefore, is a special case of Mixture SEM in which the SEM part of the model is the Latent Growth Curve Modeling.

Muthén and Shedden (1999) introduced GMM which relaxes the homogeneity assumption and allows for parameter differences across unobserved subpopulations (or latent classes). Instead of only considering individual variation around a single mean growth trajectory in LGCM, GMM can model different classes of individuals to *vary* around different mean growth trajectories. It is important to distinguish the concept of GMM from the Latent Class Growth Analysis (LCGA) model proposed by Nagin (1999), which is also a complement to HLM and LGCM. LCGA uses a multinomial modeling strategy to map group trajectories as latent classes in the data and identify homogeneous clusters of developmental trajectories (Jones, Nagin, & Roeder 2001). This "prototypal" classification recognizes fuzziness in the data, since all individuals cannot be assumed to exactly fit one and only one group trajectory (Nagin, 1999). However, LCGA treats classes as fixed; that is, individuals within a class have exactly the same developmental trajectory and the growth factor variances and covariances are assumed to be zero. Thus, LCGA is considered as a special type of the GMM (Muthen, 2004).

Unlike Latent Class Analysis and Latent Profile Analysis used for cross-sectional data, which both assume conditional independence among the observed outcomes conditional on class (i.e. the latent variable explains why the observed outcomes are related to each other), GMM does not assume uncorrelated outcomes in a given class (Muthén, 2002; Vermunt & Magidson, 2004a, 2004c). Instead, GMM allows within class random variation for the growth factors and the correlation between outcomes at all time points (Muthén, 2002). There have been some arguments about whether within-class variation should be allowed (Nagin, 2005; Muthén, 2006a). Compared to LCGA, GMM usually generate fewer classes for the same data (Nagin, 2005; Bauer & Curran, 2004). However, Nagin (2005) raised the issue of possible "group cross-overs" due to the allowed within-class variation. Muthén (2006a), on the other hand, argues that GMM can represent the data more realistically. It goes beyond the focus of this dissertation to explicate on this controversy, however, the implication for researchers is that potential importance of within-class variation is also an empirical question and worth researchers' exploration (Connell & Frye, 2006).

The GMM equations and the corresponding matrix forms (i.e. equations 3a-3j) for a linear growth mixture model with covariates predicting both the growth factors and latent class variable are the summary of the works by Muthén and Shedden (1999), Jedidi et al. (1997), Vermunt & Magidson (2005), and Bollen and Curan (2006), Palardy and Vermunt (2007) as following:

Measurement: $$Y_{it} = \alpha_i + \lambda_t \beta_i + \varepsilon_{it} \qquad \varepsilon_{it} \sim N(0, \sigma^2) \qquad (3a)$$

Structural:
$$\alpha_i = \sum_{k=1}^{K} \mu_{\alpha k} c_{ik} + \sum_{q=1}^{Q} \gamma_{\alpha q} X_{qi} + \zeta_{\alpha i} \qquad (3b)$$

$$\zeta_i \sim N(0, T_\zeta)$$

$$\beta_i = \sum_{k=1}^{K} \mu_{\beta k} c_{ik} + \sum_{q=1}^{Q} \gamma_{\beta q} X_{qi} + \zeta_{\beta i} \qquad (3c)$$

$$\log it[P(c_{ik} = 1)] = \lambda_{0k} + \sum_{q=1}^{Q} \lambda_{qk} X_{qi} \qquad (3d)$$

Measurement:
$$Y_i = \Lambda \eta_i + \varepsilon_i \qquad (3e)$$

$$\begin{bmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{iT} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & T-1 \end{bmatrix} \begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} + \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{iT} \end{bmatrix} \qquad (3f)$$

$$\mathrm{Y_i} \quad = \quad \Lambda_\mathrm{i} \qquad \eta_\mathrm{i} \quad + \quad \varepsilon_\mathrm{i}$$

Structural:
$$\eta_i = \mathrm{A} c_i + \Gamma X_i + \zeta_i \qquad (3g)$$

$$\begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} = \begin{bmatrix} \mu_{\alpha 1} & \mu_{\alpha 2} & \cdots & \mu_{\alpha k} \\ \mu_{\beta 1} & \mu_{\beta 2} & \cdots & \mu_{\beta k} \end{bmatrix} \begin{bmatrix} c_{i1} \\ c_{i2} \\ \vdots \\ c_{ik} \end{bmatrix} + \begin{bmatrix} \gamma_{\alpha 1} & \gamma_{\alpha 2} & \cdots & \gamma_{\alpha Q} \\ \gamma_{\beta 1} & \gamma_{\beta 2} & \cdots & \gamma_{\beta Q} \end{bmatrix} \begin{bmatrix} X_{1i} \\ X_{2i} \\ \vdots \\ X_{Qi} \end{bmatrix} + \begin{bmatrix} \zeta_{\alpha i} \\ \zeta_{\beta i} \end{bmatrix} \qquad (3h)$$

$$\eta_i \quad = \quad \mathrm{A} \quad c_i \quad + \quad \Gamma \quad X_i \quad + \quad \zeta_i$$

Latent Class:
$$log\, it(w_i) = \mathrm{K} + \Lambda X_i \qquad (3i)$$

$$\begin{bmatrix} log(\frac{w_{i1}}{w_{ik}}) \\ log(\frac{w_{i2}}{w_{ik}}) \\ \vdots \\ log(\frac{w_{i.k-1}}{w_{ik}}) \end{bmatrix} = \begin{bmatrix} \lambda_{01} \\ \lambda_{02} \\ \vdots \\ \lambda_{0.k-1} \end{bmatrix} + \begin{bmatrix} \lambda_{11} & \lambda_{11} & \cdots & \lambda_{1k} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ \lambda_{Q-1.1} & \lambda_{Q-1.2} & \cdots & \lambda_{Q-1.k-1} \end{bmatrix} \begin{bmatrix} X_{1i} \\ X_{2i} \\ \vdots \\ X_{Qi} \end{bmatrix} \qquad (3j)$$

$$\log it(w_i) \quad = \quad \mathrm{K} \quad + \qquad\qquad \Lambda \qquad\qquad X_i$$

where the measurement model (i.e. equations 3a, 3e, 3f) remains the same as in LGCM, while in the structural part of the model growth factors (i.e. equations 3b, 3c, 3g and 3h) are related to the categorical latent variable *c* (i.e., c_{ik} symbolizes *k* indicator variables, where $c_{ik} = 1$ if individual *i* belongs to class *k* and is zero otherwise, *k*=1, 2, 3, …, *k* and *k* is the total number of mixture components or latent classes) given the observed covariate *X*. The matrix A in equation 3g and 3h contains the intercept parameters for each *c* class, and the other parameters and assumptions are the same as previously stated for LGCM equations. Equations 3d, 3i and 3j show how the categorical latent variable *c* relates to covariate *X* through a multinomial logit regression model for unordered polytomous response using class *k* as the reference category, defining $w_{ik} = P(c_{ik} = 1 \mid X_i)$ and $w_i = (w_{i1}, w_{i2}, \cdots, w_{ik})'$.

Model Estimation

The density of Y_i, $f(y_i; \theta)$, is a mixture or weighted sum of *k* class-specific densities $f(y_i; \theta_k)$ as follows:

$$f(Y_i; X_i, w, \theta) = \sum_{i=1}^{g} w_k(X_i) f_k(Y_i; \theta_k) \qquad (3k)$$

where $0 < w_k < 1$ and $\sum_{k=1}^{K} w_k = 1$, θ_k contains the vector of unknown model parameters for class *k*, and $w_k(X_i)$ is the probability of individual *i* belong to class *k* given the corresponding covariate value X_i.

Define $z_k^{'} = \begin{bmatrix} Y_k^{'} & X_k^{'} \end{bmatrix}$ as the joint vector of observed indicators conditional on membership in class *k*, and assume that within each class the distribution of the observed variables z_k is multivariate normal. The unconditional distribution of z is then a finite mixture of z_k. The likelihood function for a sample $(z_1, z_2 \cdots z_N)^{'}$ randomly drawn from the mixture is:

$$L = \prod_{j=1}^{N} \left\{ \sum_{k=1}^{K} w_{ik} (2\pi)^{-\frac{p+q}{2}} |\Sigma_k|^{-\frac{1}{2}} \exp\left[-\frac{1}{2} (z_i - \mu_k)' [\Sigma_k]^{-1} (z_i - \mu_k) \right] \right\} \tag{3l}$$

where $i = 1,2,\cdots,N$, w_{ik} is the probability that the individual *i* belong to class *k*, μ_k is the class specific mean vector and Σ_k the class-specific covariance matrix, and *p* and *q* are the numbers of indicators for the endogenous and exogenous constructs.

For an observation in group *k*, the model implied mean vector and covariance matrix can be expressed as follows:

Mean Vector:
$$\mu^k(\theta) = \begin{bmatrix} \Lambda^k (\mu_\eta^k + \Gamma^k \mu_x^k) \\ \mu_x^k \end{bmatrix} \tag{3m}$$

Covariance Matrix:
$$\sum^k(\theta) = \begin{bmatrix} \Lambda^k (\Gamma^k \Sigma_{xx}^k \Gamma^{k\prime} + \psi^k) \Lambda^{k\prime} + \Sigma_{\varepsilon\varepsilon}^k & \Lambda^k \Gamma^k \Sigma_{xx}^k \\ \Sigma_{xx}^g \Gamma^{k\prime} \Lambda^{g\prime} & \Sigma_{xx}^k \end{bmatrix} \tag{3n}$$

where the superscript *k* is attached to each variable and parameter to show the class it belongs to, θ is the vector of the unknown model parameters for group *k*, $\Lambda^k(\mu_\eta^k + \Gamma^k \mu_x^k)$ is the model implied mean vector for *Y*, μ_x^k is the vector of means for the covariates X^k, and Σ_{xx}^k is the population covariance matrix of the X^k. Each latent group in the analysis has the similar model implied matrices. In an unconditional linear model, Σ_{xx}^k and Γ^k will be set to zero.

Model estimation for GMM is usually done using the EM algorithm or some modifications of it to maximize the likelihood function (Muthén & Sheddan, 1999; Jedidi et al., 1997; Vermunt & Magidson, 2005). This estimation method was explained in details by Dempster, Laird, and Rubin (1977). Since its advent, maximum likelihood (ML) is the most commonly used approach to fit FMM (McLachlan & Peel, 2000). By treating the continuous latent variables α_i and β_i and the categorical latent variable c_i as missing data, the E (expectation) and M (maximization) steps proceeds iteratively to maximize the *L* function in equation 3l with respect to the free parameters given in the observed sample data and a specified number of classes *k*, while also taking into account the constraints on *w* and $|\Sigma_k| > 0$ for all *k* (Jedidi et al., 1997).

Once convergence is reached, the algorithm provides estimates of the model parameters and their asymptotic covariances (McLachlan, 1992). The posterior probability of individual *i* belonging to class *k* can be calculated using the parameter estimates using Bayes' rule:

Posterior Probability:
$$P_{ik} = \frac{\hat{w}_k(X_i) f_k(Y_i; X_i, \hat{\mu}_k, \hat{\Sigma}_k)}{\sum_{h=1}^{k} \hat{w}_h(X_i) f_h(Y_i; X_i, \hat{\mu}_h, \hat{\Sigma}_h)} \quad (3o)$$

where P_{ik} is the posterior probability. Each individual is then assigned to the class with the highest posterior probability.

Model Selection

Two issues arise in the model selection process for GMM: 1) the decision about the number of classes; and 2) the form of the model given the number of classes (Jedidi et al., 1997; Vermunt & Magidson, 2002).

The form of the model can be examined using the standard likelihood-ratio test (LRT) for nested models when the number of classes has been determined (Vermunt & Magidson, 2002). An example of such nested models would be for a LCGA model and a GMM model fitted to the same set of data with the same number of latent classes, with the former having a restricted covariance matrix compared to the latter. However, LRT cannot be used for class number determination, because LRT is not asymptotically distributed as a chi-square distribution between the likelihood values of models with *k* versus *k-1* number of classes (McLachlan & Peel, 2000).

Testing for the number of components is of great theoretical and practical importance, yet it is very difficult problem and has not yet been completely resolved (McLachlan & Peel, 2000). A combination of criteria has been recommended to guide applied researchers in selecting the optimal number of classes. Nevertheless, Muthén (2003) emphasized the importance of checking on substantive knowledge before using any of these statistical criteria. The fit statistics can be grouped into four categories (Tofighi &

Enders, 2008; Vermunt & Magidson, 2002; Henson, Reise, & Kim, 2007): information-based criteria, nested model likelihood ratio tests, goodness of fit measures, and classification-based statistics. Researchers (Celeux & Soromenho, 1996; Tofighi & Enders, 2008; Nylund, Asparouhov & Muthén, 2008; Henson et al., 2007) have studied the performance of these fit statistics in the context of latent variable mixture modeling and following information about the use of these statistics is summarized from these articles.

Information Criterion (IC) indices are based on the log likelihood value of a fitted model and the penalty on model complexity such as the number of model parameters and/or sample size. The following ICs have been studied: Bayesian information criterion (BIC), a sample size adjusted BIC (SABIC), Akaike's Information Criteria (AIC), the consistent AIC (CAIC), and a sample size adjusted CAIC (SACAIC). IC indices are penalized log likelihood criteria (McLachlan & Peel, 2000), which means that as log likelihood increases with the addition of a component to a mixture model, the decreased -2 log likelihood is penalized by adding a term related to the number of parameters in the model (i.e., $BIC = -2\log L + p\log(n)$, where n is the sample size, and p is the number of estimated parameters, L is the likelihood and $0 < L \leq 1$, larger values of L yield smaller non-negative values of $-2\log L$, $p\log(n)$ is the penalty term for model complexity). A lower IC value favors model with higher likelihood value and fewer parameters, and lower IC values indicates better model fit. Because of the different penalty functions in the indices, it is possible that different ICs may favor different class solution as the best model.

The nested model likelihood ratio tests include: the Lo-Mendell-Rubin adjusted likelihood ratio test (LMR), the adjusted Lo-Mendell-Rubin likelihood ratio test (ALMR), and the bootstrap likelihood ratio test (BLRT). All these statistics were developed using

LRT and they are testing the null hypothesis that the restricted model with *k-1* classes fits the data as well as the full model with *k* classes. A small probability value of these tests indicates that the *k-1* class model should be rejected in favor of the *k* class model.

The classification-based statistics include: CLC (classification likelihood information criterion), ICL–BIC (integrated classification likelihood), normalized entropy criterion (NEC), and entropy. These statistics are calculated based on the classification of individuals and the separation of latent classes. Smaller values of CLC, ICL-BIC, and NEC, and larger values of entropy indicate a more unambiguous classification.

The goodness of fit statistics is the Multivariate Skewness and Kurtosis Tests (MKT & MKT). These statistics compare the *k*-class model-implied multivariate skewness and kurtosis values to those from the sample data. Small probability values would indicate that the *k* class model differs from the sample distribution and does not fit the data. Large probability values would then indicate the model adequately fits the data.

Of all these fit statistics, Tofighi and Enders (2008) found that the SABIC and LMR were useful in enumerating the correct number of classes, although they did not evaluate the performance of BLRT due to software limitation. Nylund et al. (2006) found that BLRT was the best among the statistics they examined (including LMR), following by BIC and then SABIC. However, they also pointed out compared to LMR, BLRT has limitations such as requiring long computation time, dependence on distributional and model assumptions, and inability to accommodate complex survey data. Therefore, they recommended using BIC and LMR to narrow the solutions to a few plausible models first and then requesting BLRT for these models to select the best model. Neither Tofighi and Enders (2008) nor Nylund et al. (2006) studied the classification-based statistics. Henson

et al. (2007) studied all four types of fit statistics and found that ABIC performed best. They also found that except for entropy, the classification statistics performed well but not as good as the information-based criteria or the nested model likelihood ratio tests, contrary to what Celeux & Soromenho (1996) found for NEC. They did not find the MST or MKT useful for their study model, whereas Tofighi and Enders (2008) speculated that the performance of MST and MKT is model-specific according to their simulations. Both Tofighi and Enders (2008) and Henson et al. (2007) found that the ALMR performed similar as LMR, whereas Nylund et al. (2006) only examined LMR.

Although the types of latent variable mixture models employed in the simulation studies vary, they seem to agree on the use of SABIC, BIC, and LMR for model selection. BLRT was studied in only one simulation study yet still worth consideration due to the consistency found in three types of mixture models (Nylund et al. 2007). Last but not the least, it can never be emphasized enough that researchers must check whether the model solution makes sense according their substantive knowledge.

Limitations

GMM certainly is a very useful and promising statistical tool for studying change over time. In addition, it allows more complex ideas of development to be examined for being part of a general latent variable framework (Muthén, 2001b). However, researchers (Bauer & Curran, 2003a, 2003b, 2004; Hipp & Bauer, 2006) have pointed out that there are potential limitations of GMM that is worth applied researchers' attention and consideration.

The first limitation is estimation difficulty, namely, the mixture model's susceptibility to local optimal solutions and model non-convergence problems (McLachlan

& Peel, 2000; Muthén, 2001b; Bollen & Curran, 2006). It is recommended that multiple sets of starting values be used to avoid theses irregularities on the likelihood surface and to discriminate local optima from the global optimum. Hipp and Bauer (2006) conducted both empirical and Monte Carlo studies to examine the consequence of failing to consider the possible presence of local optima in both GMM and LCGA models. They found that the percentage of starts converged on a solution declined as model complexity increased (i.e. model with more classes, permitting random effects within classes). Despite the previous ambiguous recommendations to vary start values, they give the following recommendations according to their study results: 1) at least 50 to 100 sets of starting values will be needed and this is especially necessary for more complex models; 2) compare the substantive results of the key solutions obtained to see whether there are similar results arising, and determine the robustness of latent classes by running models with more or fewer classes; 3) be wary that the optimal solution should occur more frequently if the model is correctly specified and vice versa; and 4) override the default start values in the computer program to specify a larger range of start values.

The second limitation, which was highlighted by Bauer and Curran (2003a, 2004), is that when the multivariate normality assumption for the repeated measures in a one-group LGCM is violated, latent trajectory classes can be estimated even in the absence of population heterogeneity. Their basic argument is that GMM has two distinct functions: the first is to identify qualitatively distinct classes of individuals with heterogeneous patterns of change over time, and the second is to approximate intractable or complex distributions with a small number of simpler component distributions. It is not easy to distinguish between the two functions analytically, nor do the fit statistics provide relevant

information. Their simulation study showed that for data with nonnormal distribution from the same population, the two-class model fit the data better than the single-class model, however, the overextracted model resulted in largely uninterpretable within class parameter estimates. The problem becomes more complicated when there is a mixture of nonnormal distributions. They cautioned applied researchers to validate the conclusion of multiple latent classes by substantive theory, check on the sources of nonnormality, and be mindful that alternative explanation for the same result exists (Bauer & Curran, 2003b, 2004).

The third limitation is that model misspecification for the latent classes can lead to wrong decision on the number of classes. Bauer and Curran (2004) studied two model misspecification situations by simulation: one was to fit a LCGA model to data generated using GMM one-class model, which led to overextraction of latent classes; the other was to fit a linear GMM model to data generated by a non-linear one class latent growth curve model and the model fit statistics showed preference for a two-class model.

A fourth issue with GMM is the inclusion of covariate in the prediction of class variable and growth factors. Although in both LGCM and GMM, it is recommended to include covariates to predict growth factors and/or the latent class variable (Muthén, 2004; Bollen & Curran, 2006), applied research and simulation studies have shown that the inclusion of covariate is detrimental to class enumeration (Chen & Willson, 2006; Tofighi & Enders, 2008). Tofighi and Enders (2008) found that the benefits of inclusion of covariates for better class enumeration was evident at extremely large sample size (i.e., N=2000) while at smaller sample sizes including covariates caused the lost of power for class enumeration. Chen and Willson (2006) examined the influence of the inclusion of an

exogenous covariate in conditional and unconditional GMM, and found that the latent class model with covariate resulted in losing power and inflating Type I error rate when comparing to the known class model. The power decreased and Type I error rate was inflated as sample size decreases and the magnitude of covariance between the covariate and latent class variable becomes smaller. Muthén (2006b) recently proposed to decide the number of classes first without including covariate in the analysis and then include the covariate in follow-up analysis.

Multilevel Growth Mixture Model

The concept of Multilevel Growth Mixture Model (MGMM) is a relatively new modeling idea (Muthén, 2004). Longitudinal data are often collected through cluster sampling and this gives rise to the multilevel data with repeated measures nested within individuals and individuals nested within organizations. Asparouhov and Muthén (2008) described MGMM, which is a statistical model integrating multilevel models, finite mixture models, and structural equation models. They also described several more complicated MGMM model including the model with between level classification. Palardy and Vermunt (2007) extended the simple MGMM (Muthén, 2004) which classifies the individuals but not the organizations into a more complex MGMM that can classify either the within (i.e., individual) or between (i.e. organization) or both levels of units. Such kind of modeling allows researchers to study the associations between organizational characteristics and individual growth patterns. This modeling framework is complicated though, as Palardy and Vermunt (2007) proposed, because it contains three possible configurations at both the within and the between levels and yields nine possible MGMM models.

As mentioned earlier, the LGCM and the LCGA models can be viewed as special cases of GMM, with LGCM modeling the random effects of the growth factors but only assuming one single class in the population, whereas the LCGA assuming multiple classes in the population without modeling the random variation for growth factors. GMM is a more general model and can be viewed as a combination of LGCM and LCGA because it models both random effects for growth factors and assuming the existence of heterogeneous classes. If there are three alternative approaches (i.e., LGCM, LCGA, and GMM) to account for the heterogeneity at both the individual level growth trajectories and the organizational level growth trajectories, researchers then have nine approaches to model the multilevel longitudinal data depending on whether they assume heterogeneity and/or model the random effects at each level. The focus of this dissertation is on MGMM with GMM model at the individual level and LGCM at the organizational level. Palardy and Vermunt (2007) pointed out that some of the approaches are more useful than others, nevertheless, the most general formulation of the MGMM which use GMM at both levels in the model, is introduced below for heuristic purposes. Notice that the matrix terms are not listed here and they would look very similar as the ones shown in equations 3e-3j.

Measurement:
$$Y_{ijt} = \alpha_{ij} + \lambda_t \beta_{ij} + \varepsilon_{ijt} \qquad \varepsilon_{ijt} \sim N(0, \sigma^2) \tag{4a}$$

Within level structural:
$$\alpha_{ij} = \sum_{k=1}^{K} \mu_{\alpha jk} c_{ijk} + \sum_{q=1}^{Q} \gamma_{\alpha qj} X_{qij} + \zeta_{\alpha ij} \tag{4b}$$
$$\zeta_{ij} \sim N(0, T_\zeta)$$
$$\beta_{ij} = \sum_{k=1}^{K} \mu_{\beta jk} c_{ijk} + \sum_{q=1}^{Q} \gamma_{\beta qj} X_{qij} + \zeta_{\beta ij} \tag{4c}$$

Within level latent class: $$\log it[P(c_{ijk}=1)] = \lambda_{0k} + \sum_{q=1}^{Q} \lambda_{qk} X_{qij} \quad (4d)$$

Between level $$\mu_{\alpha j} = \sum_{l=1}^{L} \nu_{0l} d_{jl} + \sum_{p=1}^{P} \nu_{0p} W_{pj} + \rho_{0j} \quad (4e)$$

$$\rho_j \sim N(0, T_\rho)$$

$$\mu_{\beta j} = \sum_{l=1}^{L} \nu_{1l} d_{jl} + \sum_{p=1}^{P} \nu_{1p} W_{pj} + \rho_{1j} \quad (4f)$$

Between level latent class: $$\log it[P(d_{jl}=1)] = \delta_{0l} + \sum_{p=1}^{P} \delta_{ql} W_{pj} \quad (4g)$$

where Equations 4a-4d show the within level of MGMM, which remains largely the same as equations 3a-3d for GMM except for subscript *j* is added to stand for the number of clusters at the organizational level. Latent classes can be drawn from either the individual or organizational level, or both. Equations 4e-4g describes the between level of MGMM, which is another set of GMM models but with different symbols. *L* (*l* =1, 2, 3, …, *l*) is used to denote the number of classes at the organizational level, so as to distinguish from the individual level number of classes *K (k)*. d_{jl}, functioning similar as c_{ijk} but for the organizational level, symbolizes the indicator variables equal to 1 if an organization *j* belongs to between-level latent class *l* and 0 otherwise. ν_{0l} and ν_{1l} are the mean intercept and slope for class *l*, whereas ν_{0p} and ν_{1p} are the regression coefficients between organizational level covariate W_{pj} with a total number of *p* (*p* =1, 2, 3, …, *p*). Equation 4g describes using between level covariate to predict the organizational class membership.

The Necessity of Studying MGMM

The combination of both continuous and categorical latent variables makes GMM a flexible analysis framework (Muthén & Muthén, 2006-2007). However, when researchers

analyze their data using GMM, some of them may assume that the participants are independent from each other even though it may not always be true. For example, in educational setting, the data structure is very likely to contain three or more levels (e.g., repeated measures nested within students and students nested within schools). Nevertheless, some researchers analyzed their data using GMM without considering the higher level nesting structure (e.g., schools) and assuming the independence between students (e.g., D'Angiulli, Siegel, & Maggi, 2004; Boscardin, Muthén, Francis, & Baker, 2008).

The reasons for ignoring a level in analysis have been reviewed, including: to reduce the complications in data analysis (Wampold & Serlin, 2000; Meyers & Beretvas, 2006), to compensate the lack of identifiers on all possible levels of nesting in data (Moerbeek, 2004), and to reduce the difficulty in achieving convergence in model estimation (Van Landeghem, Fraine & Damme, 2005). Although some of these reasons may be justifiable, it is still important to examine the impact of ignoring a level of nesting structure in MGMM and to provide researchers with recommendations when ignoring a level of nesting structure is not avoidable.

The impact of ignoring a level in multilevel model has been studied and discussed (Wampold & Serlin, 2000; Moerbeek, 2004; Van Landeghem et al., 2005; Meyers & Beretvas, 2006; Luo & Kwok, 2006) and mixed findings have been reported. Some common findings are that the variance of the ignored level was redistributed to adjacent levels or non-ignored cross-classified factors and the fixed effect estimates are not affected (Moerbeek, 2004; Meyers & Beretvas, 2006; Luo & Kwok, 2006). However, there were reports of both inflation of Type I error rate due to underestimation of standard errors

(Wampold & Serlin, 2000; Meyers & Beretvas, 2006) and decreased power resulting from inflation of Type II error rate due to overestimation in standard errors (Moerbeek, 2004).

The topic of Multilevel Growth Mixture Models is relatively new (Asparouhov & Muthén, 2008, Palardy & Vermunt, 2007) and the impact of ignoring a level of nesting structure in MGMM has not yet been well examined. The purpose of this dissertation is to examine the impact of ignoring a higher nesting structure in MGMM on the accuracy of classification of individuals and the accuracy as well as the test of significance (i.e., Type I error rate and statistical power) of the parameter estimates for the model of each subpopulation.

Two studies were conducted. In Study One, the impact of misspecifying the multilevel mixture model, a model similar to MGMM, was investigated by ignoring a level of nesting structure in cross-sectional data. In Study Two, longitudinal clustered data (e.g. repeated measures nested within students and students nested within schools) was analyzed with misspecifications such as ignoring the highest level (school level) of the nesting structure. The impact of misspecification on the accuracy of classification, and the estimation and statistical inference of the parameters within each latent class was investigated.

CHAPTER II

STUDY ONE

Because FMM is the foundation of GMM, as a preliminary step, Study One examined the impact of ignoring a higher nesting level in multilevel mixture model on the accuracy of classification of individuals and the accuracy as well as the test of significance (i.e., Type I error rate and statistical power) of the parameter estimates for the model of each subpopulation. A two-level data structure (e.g., students nested within schools) was considered. Two latent classes with known group memberships were generated and then analyzed by the true (mixture modeling considering the higher level structure) and false (mixture modeling ignoring the higher level structure) models. Study One was composed of two simulations, namely, Simulations 1A and 1B. In Simulation 1A, the two latent classes were balanced in both sizes and variances, whereas in Simulation 1B the two latent classes were unbalanced in sizes and variances. The method of each simulation is first described, followed by the results and discussion.

Simulation 1A

Method

Data Generation

In Simulation 1A, data with two known subpopulations under a two-level model was first generated with equal population sizes. Then, the data was analyzed as a two-level model (i.e., true model) using multilevel mixture modeling and as a single-level model (i.e., false model) using mixture modeling. Mixture modeling is referred as Latent Class Analysis (LCA) in the following presentation. The two-level model for data generation is shown below:

Level 1: $$Y_{ij} = \beta_{0j} + \beta_{1j} Subpopulation_{ij} + e_{ij} \quad (5a)$$

$$\text{with } e_{ij} \sim N(o, \sigma^2) \quad (5b)$$

Level 2: $$\beta_{0j} = \gamma_{00} + \mu_{0j} \quad (5c)$$

$$\beta_{1j} = \gamma_{01} \quad (5d)$$

$$\text{with } \mu_{0j} \sim N(o, \tau_{00}) \quad (5e)$$

where $subpopulation_{ij}$ was a dichotomized variable with 0 and 1 representing two different subpopulations.

Suppose level 1 is the student level and level 2 is the school level. There were 40 schools, and within each school there were 20 students. The number of students in each subpopulation was 400, since the mixing proportion was set to be 50% vs. 50%. Within each school, there were 20 students coming from 2 subpopulations, 10 at risk vs. 10 non-at-risk. Altogether there were 800 students within each replication for data generation. The number of higher level units was set to be 40 given that the recommended minimum number of higher level units for multilevel mixture models is 30 (Bengt Muthén, 2005; Linda Muthén, 2003).

In this two-level model, a total of 4 parameters needed to be specified: two fixed effect coefficients (i.e., γ_{00} & γ_{01}) and two variances of the random effects (i.e., σ^2 & τ_{00}). Before specifying the population parameters in the conditional model, a random intercept model in which there are no subpopulations is presented as follows:

Level 1: $$Y_{ij} = \beta_{0j}^{*} + e_{ij} \quad (6a)$$

$$\text{with } e_{ij}^{*} \sim N(o, \sigma^{2^{*}}) \quad (6b)$$

Level 2: $$\beta_{0j}^{*} = \gamma_{00}^{*} + \mu_{0j}^{*} \quad (6c)$$

$$\text{with } \mu_{0j}^{*} \sim N(0, \tau_{00}^{*}) \quad (6d)$$

The variance of the random effect at level 1 was specified following Raudenbush and Liu's (2001) criteria, namely, $\sigma^{2^{*}} = 1$.

For τ_{00}^{*}, the intra-class correlation (ICC) formula $ICC = \tau_{00} / (\sigma^2 + \tau_{00})$ was used to obtain the values corresponding to small and medium effect size. By fixing ICC equal to 0.10 as a small ICC which is very common for studies in education (Hox, 2002) and 0.20 as a medium ICC, the values for a small τ_{00}^{*} (0.111) and a medium τ_{00}^{*} (0.250) was obtained.

According to Sinjders & Bosker (1999), adding a predictor (i.e. $\text{subpopulation}_{ij}$) at level 1 only contributes to the variance of the level-1 random errors, but does not contribute to between level variance. The formulas for calculating the within- and between-variances when there is multilevel structure in the data are $\sigma^2_{within} = \sigma^2$ and $\sigma^2_{between} = \tau_{00} + (\sigma^2 / n)$, where n is the number of students per school.

Using these formulae for calculation, a small (0.161) and a medium (0.300) $\sigma^2_{between}$ for the random intercept model was obtained. After adding $\text{subpopulation}_{ij}$ as a predictor at level one, β_{1j} was actually the difference between the two subpopulations within each school (cluster), and γ_{01} was the average difference between these two subpopulations

across all clusters. The effect size R^2 was used to characterize the difference between the two subpopulations with small, medium, and large effect sizes being 0.1, 0.3, and 0.5 according to Cohen (1988, 1992). $R^2 = 0.5$ meant that 50% of the variance between the two subpopulations could be explained by their group membership. Therefore, the larger the R^2, the larger the difference between the two subpopulations.

Using the R^2 information, small, medium, and large β_{1j} values could be calculated and was 0.632, 1.095, and 1.414 respectively. The corresponding σ^2 in the conditional model for small, medium and high levels of group difference was 0.9, 0.7, and 0.5. τ_{00} for the conditional model could be solved using equation $\sigma^2{}_{between} = \tau_{00} + (\sigma^2 / n) = \tau_{00}{}^{*} + (\sigma^{2*} / n)$, because ICC magnitude stayed the same across the random intercept and the conditional models. For ICC = .1, τ_{00} was 0.116, 0.126, and 0.136 for small, medium and large effect sizes respectively; for ICC = .2, τ_{00} was 0.255, 0.265, and 0.275 for small, medium and large effect sizes respectively.

After fixing γ_{00} to 1, the mean for Subpopulation A and the mean for Subpopulation B were calculated using equation (5a). The mean of Subpopulation A was 1 in all conditions, and the means for Subpopulation B were 1.632, 2.095, and 2.414 at different levels of R^2.

In summary, by specifying R^2 (0.1, 0.3, & 0.5) and ICC (0.1 & 0.2) values, and also set $\delta^2 = 1$, $\gamma_{00} = 1$, the population parameter values for the other fixed effect coefficient (i.e., γ_{01}) and the two variances of the random effects (i.e., σ^2 & τ_{00}) were obtained.

The simulation used a 3 (effect sizes—amount of variance explained by group membership) * 2 (magnitude of ICC) factorial design to generate the data. A total of 500 replications were generated for each condition using SAS 9.1, yielding a total of 3000 datasets. Each dataset was then analyzed by a true model (LCA considering the higher/cluster level, type = twolevel mixture) and a false model (LCA ignoring the higher/cluster level, type = mixture) using Mplus 4.2 Mixture routine (Muthén & Muthén, 2006-2007).

Analysis

For each condition, valid replications for data analysis were selected because among the replications with converged results, there were latent classes with very few students (i.e. 1 or 2). A valid replication was defined as one of the two subpopulations (or classes) with class size at least equal to or larger than 6% of the total sample size (i.e. 48 out of 800). This 6% criterion was based on the average percentage of sample size for the smallest class in published studies using LCA found in PsycINFO database.

The accuracy of classification of individuals, and the accuracy as well as the test of significance (i.e., Type I error rate and statistical power) of the parameter estimates of the model for each subpopulation were then evaluated.

Hit rate is the percentage of at-risk/non-at-risk students correctly classified as at-risk/non-at risk. The true and false models were evaluated by comparing the hit rate difference between the two models.

The group mean parameter estimates from the true and false models were summarized across the valid replications for each of the 6 conditions. The relative bias (RB) for each parameter estimate was calculated using the following equation:

$$B(\hat{\theta}) = \frac{\overline{\hat{\theta}}_{est} - \theta_{pop}}{\theta_{pop}} \quad (7)$$

where $\overline{\hat{\theta}}_{est}$ is the mean of a parameter estimate across the valid replications and θ_{pop} is the true parameter value. RB equal to zero indicates an unbiased estimate of the parameter. A negative RB indicates an underestimation of the parameter (i.e. the estimated value is smaller than the true parameter value), whereas a positive RB indicates an overestimation of the parameter (i.e. the estimated value is larger than the true parameter value). The cutoff value of 0.05 recommended by Hoogland and Boomsma (1998) was used for acceptable RB of parameter estimates.

The RB of estimated standard errors was computed using the following equation:

$$B(\hat{S}_{\theta}) = \frac{\overline{\hat{S}}_{\hat{\theta}\_False} - \overline{\hat{S}}_{\hat{\theta}\_True}}{\overline{\hat{S}}_{\hat{\theta}\_True}} \quad (8)$$

where $\overline{\hat{S}}_{\hat{\theta}\_False}$ is the mean of the estimated standard errors of the group mean parameter estimate across the valid replications in the false model, and $\overline{\hat{S}}_{\hat{\theta}\_True}$ is the standard deviation of the parameter estimate across the valid replications in the true model within a particular design condition. The standard deviation was obtained after fitting the correctly specified model to the data (i.e., the model considering the higher level nesting structure), and thus represents the "true" sampling variation, or standard error, that would have been

achieved had the model been properly specified. Hoogland and Boomsma (1998) recommended a cutoff value of 0.10 for acceptable relative bias of estimated standard errors.

Analyses of variance (ANOVAs) were conducted to determine the contribution of the two design factors (i.e., R^2 & ICC) and their interaction effect, with η^2 (i.e., $\eta^2 = SS_{Effect} / SS_{Total}$) as the effect size indicator. η^2 instead of the significance test was used because with the large number of records, the sum of square error was substantially reduced and any tiny effect could be detected as significant using the *F* test. Therefore, $\eta^2 \geq .01$ was adopted as the effect size indicator to filter out the effects trivial in magnitude and to evaluate the impact of design factors.

Results

Hit Rate

Table 1 presents the number of valid replications in Simulation 1A and the average hit rate under true and false models across valid replications. The results show that as group difference increased, the hit rate increased for both true and false model. Besides, within the same design condition, the hit rate under true model is always higher than that under false model. As ICC increased, the difference in hit rate between true and false models increased.

ANOVA results indicate that only R^2 (F (2, 2257) = 12217.44, p<.001, η^2 = .91) had substantial impact on the true model hit rate, which increased as R^2 increased. On the other hand, for the false model, both R^2 (F (2, 2257) = 5904.61, p<.001, η^2 = .83) and ICC (F (1, 2257) = 142.551, p<.001, η^2 = .01) had impact. The hit rate under false model increased as R^2 increased, but decreased as ICC increased. For the difference in hit rate

between true and false models, there was an interaction effect between R^2 and ICC (F (2, 2257) = 24.55, p<.001, η^2 = .02), which was shown in Figure 4. As R^2 and ICC increased, the difference in hit rate between true and false models increased.

Relative Bias of Group Mean Estimates

Table 2 presents the mean relative bias (RB) of group mean estimates across valid replications under true and false models. There was an underestimate of Class 1 (the smaller mean) mean and an overestimation of Class 2 (the larger mean) mean under both true and false models when R^2 = .1. When R^2 = .3 and .5, the mean RBs under both models were close to zero, except for the mean RB for Class 1 was underestimated slightly when ICC = 0.2.

ANOVA results showed that only R^2 (*F*s (2, 2257) = 323.481 and 606.988, *p*s<.001, η^2 s = .22 and .35 for Class 1 and Class 2 respectively) had substantial impact on the relative bias of group mean estimates under true model, which decreased as R^2 increased. Similar results were found for false model (*F*s (2, 2257) = 366.814 and 681.945, *p*s<.001, η^2 s = .24 and .38 for Class 1 and Class 2 respectively).

Relative Bias of Variance Estimates

Table 3 presents the mean RBs of variance estimates of the true and false model. For the true model, the mean RB of most level-1 and level-2 variance estimates were within $\pm 10\%$, where as for the false model, there was a trend of overestimation in level-1 variance estimates.

ANOVA results indicated that R^2 (*F* (2, 2257) = 506.515, *p*<.001, η^2 = .31) had substantial impact on the RB of level-1 variance estimates under true model, and ICC (*F* (1, 2257) = 32.515, *p*<.001, η^2 = .01) had an impact on level-2 variance estimates under

true model. For the false model, there was an interaction effect between R^2 and ICC (F (2, 2257) = 43.726, p<.001, η^2 = .014), which was shown in Figure 5.

Relative Bias of Standard Errors of Group Mean Estimates

Table 4 presents the mean RBs of standard errors for group mean estimates under the false model. There was an inflation of standard errors for group mean estimates under the false model. ANOVA results show that R^2 (*F*s (2, 2257) = 10.017 and 13.931, *p*s<.001, η^2 s = .009 and .012 for Class 1 and Class 2 respectively) was the major source of impact when RBs of the standard errors for group mean estimates were the dependent variables.

Simulation 1B

Method

Data Generation

To extend the findings from Simulation 1A which was based on the balanced design (i.e., the two classes had exactly same number of observations and variance across clusters), Simulation 1B was conducted by taking the unbalanced sample size and variance (i.e., unequal class size for the two subpopulations) into account along with other design factors as considered in Simulation 1A. There were two imbalance types, Imbalance Type 1 and Imbalance Type 2. In Imbalance Type 1, large size was associated with large variance in Group 1 and small size was associated with small size in Group 2; in Imbalance Type 2, large size was associated with small variance in Group 1 and small size was associated with large size in Group 2. The group size and variance varied at level 1 for the two latent classes. A large group size was a group of 15 students, whereas a small group size is a group of 5 students. A larger variance group has a variance three times of the variance of the smaller variance group, so that the variance between the two latent

groups was distinguishable. Equation (9) was used to calculate the variances of each individual group based on the size of each group. The value of S_p^2, which was the pooled level 1 variance of the two latent classes, was set to be 0.9, 0.7, and 0.5 respectively because the variance accounted for by group membership was 0.1, 0.3, and 0.5 in Simulation 1A.

$$S_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \tag{9}$$

The simulation used a 3 (amount of variance explained by group membership) * 2 (magnitude of ICC) * 2 (Imbalance Type) factorial design to generate the data. A total of 500 replications were generated for each condition using SAS 9.1, yielding a total of 6000 datasets. Each dataset was then analyzed by a true model (LCA considering the higher/cluster level) and a false model (LCA ignoring the higher/cluster level) using Mplus 4.2 Mixture routine (Muthén & Muthén, 2006).

Analysis

Similar to Simulation 1A, valid replications were selected, with hit rates and relative biases of parameter estimates under the 12 conditions for both true and false models calculated and examined. Analyses of variance (ANOVAs) were conducted to determine the contribution of the design factors and all possible interactions.

Results

Hit Rate

Table 5 presents the number of valid replications for Simulation 1B and the average hit rate under true and false models. Similar to the results found in Study One, as group difference increased, the hit rate increased for both true and false models. Besides, the hit rate under true model was always higher than that under false model within the same condition. As ICC increased, the difference in hit rate between true and false models increased. In addition, Imbalance Type 2 (i.e., large variance associated with small class) always had higher hit rates than Imbalance Type1 (i.e., large variance associated with large class) when all other conditions remained the same.

ANOVA results indicated that there was an interaction effect between the magnitude of R^2 and Imbalance Type ($F(2, 3642) = 1028.61$, $p<.001$, $\eta^2 = .15$ for true model; $F(2, 3642) = 359.02$, $p<.001$, $\eta^2 = .08$ for false model) for both the true and false models when the hit rate was the dependent variable. The hit rate increased for both Imbalance Types as R^2 increased. However, When R^2 was low, the difference between the two Imbalance Types was larger than when R^2 was high. The hit rate for Imbalance Type2 was higher than that for Imbalance Type1. Under the false model, when other conditions stay the same, hit rate was higher when the ICC value was smaller ($F(1, 3642) = 79.92$, $p<.001$, $\eta^2 = .01$).

There was an interaction effect between the magnitude of R^2 and Imbalance Type on the hit rate difference between true and false models ($F(2, 3642) = 22.56$, $p<.001$, $\eta^2 = .01$). As shown by Figure 6, the estimated mean hit rate difference between true and false models increased for both imbalance types as R^2 increased. Hit rate under true model was

higher than that under the false model. However, at higher levels of R^2, the difference in hit rate for Imbalance Type 1 is larger than that for Imbalance Type 2. Besides, when other conditions stayed the same, difference in hit rate was larger when the ICC value was larger (F (1, 3642) = 110.85, p<.001, η^2 = .03).

Relative Bias of Group Mean Estimates

Table 6 presents the mean RBs of group mean estimates under true and false models. There was bias outside the range of $\pm 10\%$ for both the true and false models. ANOVA results indicated that there was an interaction effect between R^2 and ICC (*F*s (2, 3642) = 449.637 and 92.023, *p*s<.001, η^2s = .15 and .04 for the two classes in the true model; *F*s (2, 3642) = 253.900 and 45.950, *p*s<.001, η^2s = .09 and .02 for the two classes in the false model) when the RBs of Class 1 and Class 2 were the dependent variables separately. As shown in Figures 7-10, the mean RB decreased for both Imbalance Types as R^2 increased. There were more biases under Imbalance Type 1 than Imbalance Type 2. There tended to be more biases for Class 1 (smaller mean) mean estimate than that for Class 2 (larger mean).

Relative Bias of Variance Estimates

Table 7 presents the mean RBs of variance estimates of the true and false model. Because the level-1 variances for two groups were estimated separately in both the true and the false models, there were two σ^2 s for each model. For the true model, the mean RBs for level-2 variance estimates were within or close to $\pm 10\%$, and there was no $\eta^2 \geq .01$ when RB of τ_{00} was the dependent variable. For level-1 variance, there was underestimation for σ_1^2 and overestimation for σ_2^2 under Imbalance Type 1; whereas there was less biases for Imbalance Type 2. ANOVA results indicated that there was an

interaction effect between R^2 and Imbalance Type (Fs(2, 3642) = 68.793 and 293.125, ps<.001, η^2s = .035 and .126 respectively). Figures 11 and 12 show the interaction effects.

For the false model, there was a trend of overestimation in σ_2^2 under both Imbalance Types, whereas there was both underestimation and overestimation of σ_1^2 only under Imbalance Type 1. ANOVA results indicated that there was an interaction effect between R^2 and Imbalance Type (Fs(2, 3642) = 57.494 and 34.857, ps<.001, η^2s = .027 and .012 respectively). In addition, ICC has a substantial impact on σ_2^2 overestimation (F(1, 3642) = 367.945, ps<.001, η^2s = .065). Figures 13 and 14 show the interaction effects.

Relative Bias of Standard Errors of Group Mean Estimates

Because the level-1 variances were estimated separately, there were two RBs of standard errors under each model. RBs of SE1 are for the large variance groups and RBs of SE2 are for the smaller variance group under both Imbalance Types. Again, there was a tendency of inflation of standard errors under the false model under most conditions. ANOVA results indicated that Imbalance Types (Fs (1, 3642) = 99.10 & 651.57, ps<.001, η^2s= .03 & .13) and ICC (Fs (1, 3642) = 82.85 & 60.81, ps<.001, η^2s= .02 & .01) were the two major contributing factors, although there was a slight interaction effect between them for SE1. The false model had more inflation of standard errors under Imbalance Type 1 than Imbalance Type2. Besides, within the same Imbalance Type, bias was higher at higher level of ICC.

Discussion

The purpose of this study was to investigate the impact of ignoring a higher level nesting structure in cross-sectional data. Two simulations were conducted, one including

R^2 and ICC as design factors, and the other including R^2, ICC, and Imbalance Type. The accuracy of classification of individuals, and the accuracy as well as the test of significance (i.e., Type I error rate and statistical power) of the parameter estimates of the model for each subpopulation are then examined. Multilevel data with two subpopulations were generated and then analyzed using the true model (Multilevel Mixture Model) and the false model (single-level Mixture Model). The results of each simulation are summarized and explained.

Simulation 1A

When a higher level structure in cross-sectional data is ignored, the variance at the higher level is redistributed to the lower level, thus affecting the hit rate and group mean and standard error estimates.

Hit Rate

R^2 is an important factor influencing hit rate. For both the true and false models, hit rate increases when the R^2 increases, which means that as group difference becomes larger, the classification under both models will become more accurate and this is quite reasonable.

The difference between true and false model is that for true model, ICC magnitude does not affect hit rate much within the same design. Whereas for false model, ICC magnitude affects the hit rate, and the hit rate is higher when ICC is smaller. Under the false model the level-2 variance is ignored in model estimation, and more variance is ignored at higher ICC. Obviously ignoring variance at level-2 will decrease classification accuracy, and the more variance ignored, the less accurate the classification.

Relative Bias in Group Mean Estimates

The difference in RB for group mean estimates between true and false models are all within ±5%, which indicates that the true and false models do not differ tremendously in the estimates of the group means. In other words, there was no substantial difference on the group mean estimates between the true and false models.

Relative Bias in Standard Error Estimates

There is an inflation of standard errors for group mean estimates when a higher level nesting structure is ignored. This inflation of standard errors under the false model is due to the redistribution of level-2 variance to level-1. When ICC is larger, false model has more inflation of standard errors when all other conditions stay the same.

Simulation 1B

After adding one more design factor—Imbalance Type, the findings in *Simulation 1B* related to R^2 and ICC remain consistent with findings in *Simulation 1A*. Therefore, the following discussion focuses on the influence of Imbalance Type.

Hit Rate

When all other conditions stay the same, the hit rate under Imbalance Type 2 is higher than that under Imbalance Type 1. In addition, the difference in hit rate between true and false models is smaller for Imbalance Type2, in which large group size is associated with smaller variance and small group size associated with larger variance. This means that under Imbalance Type 2, the false model's performance is relatively better than the false model under Imbalance Type 1. This result is not surprising, because when a group has smaller variance, it is easier to identify them as coming from the same group. In Imbalance Type2, when large size is associated with smaller variance, the subjects within this group have a higher chance of being classified as the same group. Compared to

Imbalance Type1, where smaller group size is associated with smaller variance, although the subjects within this group have a higher chance of being classified as the same group, they are still a smaller percentage of all subjects comparing to that in Imbalance Type2. This is why in general the Imbalance Type2 has higher hit rates than Imbalance Type1.

Relative Bias in Group Mean Estimates

In general, the RBs under Imbalance Type 2 are smaller than that under Imbalance Type 1. For the same reason mentioned before, for Imbalance Type 2, it is easier for both the true and false models to classify the subjects into the correct group, therefore resulting in more accurate estimate of the group mean. Whereas for Imbalance Type1, there are more RBs under different levels of R^2, most likely resulting from the wrong classification of subjects into wrong groups.

Relative Bias in Standard Error Estimates

When a higher level nesting structure is ignored, the standard errors of the fixed effects (i.e., the means of the two latent classes) tend to be inflated under Imbalance Type 1, but have less bias or underestimation under Imbalance Type 2. This may result from either the mis-classification of subjects, or the inflation of level-1 variance, or both.

CHAPTER III

STUDY TWO

Study Two examined the impact of ignoring a higher nesting level in multilevel growth mixture model on the accuracy of classification of individuals and the accuracy as well as the test of significance (i.e., Type I error rate and statistical power) of the parameter estimates for the model of each subpopulation. A three-level data structure (e.g., repeated measures nested within students nested within schools) was considered. Two latent classes with known group memberships were generated and then analyzed by the true (MGMM considering the higher level structure) and false (GMM ignoring the higher level structure) models. The method of the study is first described, followed by the results and discussion.

Method

Data Generation

Although no simulation studies have been done to examine the impact of ignoring a level of nesting structure in MGMM, some simulation studies related to growth mixture models have been published lately and there are some consistent findings related to several design factors. First, it was discovered that the level of class separation had a dramatic impact on the ability to correctly enumerate the number of classes (Henson, Reise, & Kim, 2007; Tofighi & Enders, 2008). If the generated classes were well-separated, the correct number of classes was easier to identify (Henson, et al., 2007; Tofighi & Enders, 2008) and the percentage of individuals correctly classified is higher (Chen, Kwok, & Luo, 2007). Second, the number of repeated measures had minor impact on class enumeration (Tofighi & Enders, 2008; Enders & Tofighi, 2008). Third, the mixing proportion of latent

classes had a substantial impact on class enumeration (Henson, et al., 2007; Tofighi & Enders, 2008; Enders & Tofighi, 2008). In unbalanced situations where one latent class had an extremely low mixing proportion (i.e. 7% in Tofighi & Enders, 2008, 10% in Henson, et al., 2007), the model was less likely to converge and the class enumeration was less accurate. Fourth, sample size influenced the performance of fit indices in class enumeration and fit indices performed better with larger sample size (Henson, et al., 2007; Tofighi & Enders, 2008). Given the previous findings and the focus of this study, mixing proportions of latent classes and sample size were included as design factors, whereas the number of repeated measures was fixed and the level of class separation was set at a relatively easy-to-separate level.

For this study, data with two known subpopulations under a three-level model were first generated. Then, the data were analyzed as a two-level model using GMM and as a three-level model using MGMM. The three-level model for data generation is shown below:

Level 1: $$Y_{tij} = \pi_{0ij} + \pi_{1ij}(Time)_{tij} + e_{tij} \quad (10a)$$

$$\text{with } e_{tij} \sim N(o, \sigma^2) \quad (10b)$$

Level 2: $$\pi_{0ij} = \beta_{00j} + \beta_{01j} Subpopulation_{ij} + r_{0ij} \quad (10c)$$

$$\pi_{1ij} = \beta_{10j} + \beta_{11j} Subpopulation_{ij} + r_{1ij} \quad (10d)$$

$$\text{with } \begin{bmatrix} r_{0ij} \\ r_{1ij} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{\pi 00} & \tau_{\pi 01} \\ \tau_{\pi 10} & \tau_{\pi 11} \end{bmatrix} \right) \quad (10e)$$

Level 3: $$\beta_{00j} = \gamma_{00} + \mu_{0j} \quad (10f)$$

$$\beta_{01j} = \gamma_{01} \quad (10g)$$

$$\beta_{10j} = \gamma_{10} \quad (10h)$$

$$\beta_{11j} = \gamma_{11} \quad (10i)$$

$$\text{with } \mu_{0j} \sim N(o, \tau_{\beta 00}) \quad (10j)$$

where subpopulation$_{ij}$ was a dichotomized variable with 0 and 1 to represent two different subpopulations.

In this three-level model, a total of 9 parameters needed to be specified: 4 fixed effect coefficients (i.e., γ_{00}, γ_{01}, γ_{10}, & γ_{11}) and 5 variances and covariance of the random effects (i.e., σ^2, $\tau_{\pi 00}$, $\tau_{\pi 01}$, $\tau_{\pi 11}$, $\tau_{\beta 00}$). The average growth models for the two subpopulations were specified as follows so that they represent "well-separated classes" following the design by Nylund, Asparouhov, & Muthén (2007), with Subpopulation A representing a slow-growing group and Subpopulation B representing a fast-growing group.

Subpopulation A (Slow Growing Class): $\hat{Y}_{tij} = 1.00 + .10 * (Time)_{tij}$ (11a)

Subpopulation B (Fast Growing Class): $\hat{Y}_{tij} = 2.50 + .50 * (Time)_{tij}$ (11b)

where γ_{00} was equal to 1, γ_{01} was equal to 1.5, γ_{10} was equal to 0.1, and γ_{11} was equal to 0.50. The intercepts for each subpopulation were 1.00 (I1), 2.50 (I2), and the slopes were .10 (S1), .50 (S2) respectively.

Following Raudenbush and Liu's (2001) criteria, the variances and covariance of the random effects were specified as follows:

$\sigma^2 = 1.0$,

$T_\pi = \begin{bmatrix} \tau_{\pi 00} & \tau_{\pi 01} \\ \tau_{\pi 10} & \tau_{\pi 11} \end{bmatrix} = \begin{bmatrix} .100 & .025 \\ .025 & .050 \end{bmatrix}$ as small T_π and

$T_\pi = \begin{bmatrix} \tau_{\pi 00} & \tau_{\pi 01} \\ \tau_{\pi 10} & \tau_{\pi 11} \end{bmatrix} = \begin{bmatrix} .200 & .050 \\ .050 & .100 \end{bmatrix}$ as medium T_π.

By doing so, the growth factor means were held constant for the two subpopulations, but the magnitude of within-class variance parameters vary, mimicking the "high separation" and "low separation" classes described in Tofighi & Enders (2008).

Equation $ICC = \tau_{\beta 00} / (\sigma^2 + \tau_{\pi 00} + \tau_{\beta 00})$ for intra-class correlation (ICC) calculation was used to obtain the values of $\tau_{\beta 00}$ corresponding to small and medium T_π. Similar as in Study One, ICC values were set to be .10 and .20. The values for a small $\tau_{\beta 00}$ (0.122) and a medium $\tau_{\beta 00}$ (0.133) for ICC = .10 and a small $\tau_{\beta 00}$ (0.275) and a medium $\tau_{\beta 00}$ (0.300) for ICC = .20 were then obtained.

Additionally, different sample sizes associated with different levels were considered (i.e., number of repeated measures, number of individuals, and number of higher level clusters (e.g., schools)). Four waves was chosen as the number of repeated measures based on two criteria. On the one hand, according to Khoo et al.'s (2006) review of the multiwave longitudinal studies published in *Developmental Psychology* in 2002, more than half (52%) of these studies collected data with three or four occasions and the mean number of waves of the other 48% of studies was 8. On the other hand, because in previous GMM simulation studies the number of repeated measures had no significant impact and 4 waves of repeated measures were used and reported in several simulation studies (Tofighi & Enders, 2008; Enders & Tofighi, 2008; Nylund, et al., 2007). The time

variable was centered so it had a mean of 0 and 1 unit between adjacent observations (i.e., $\underline{\text{Time}}_{4\text{waves}}$' = [-1.5 -.5 .5 1.5]).

For the number of higher level clusters, 30 was adopted as a small number of clusters and 50 as a large number of clusters according to Bengt (2005) and Linda Muthén's (2003) suggestions on the Mplus online discussion forum. For the number of participants nested within each cluster, based on the conditions used in past simulation studies, thirty (individuals) was the smallest sample size considered in both Keselman, Algina, Kowalchuk, & Wolfinger (1998) and Ferron, Dailey, & Yi (2002). However, from a realistic point of view, in research projects funded by national grants, sample size as large as 900 (i.e. 30 individuals nested within 30 clusters) was not easily obtained. Take a relatively funded large-scale project, Project Achieve funded by the NICHD as an example, there were 784 students nested within 36 elementary schools at the starting data collection years. On the other hand, in previous simulation study, it was found that larger sample sizes were always beneficial for class enumeration and model convergence (Henson, et al., 2007; Tofighi & Enders, 2008). It is the interest of this study to find out how MGMM would perform in relatively smaller and more realistically obtainable samples, therefore, 20 was chosen as a small cluster size and 40 as a large cluster size. Combining the cluster number (i.e., school number) and cluster size (i.e., number of students nested within each school) conditions, altogether there were four sample sizes conditions, namely, 600, 1000, 1200, and 2000.

The mixing proportions of the two subpopulations were set to be balanced or unbalanced. In the balanced situation, mixing proportion was set to be 50% and 50% for the two subpopulations. In the unbalanced situation, mixing proportion was set to be 25%

for the slow growing group and 75% for the fast growing group following Nylund et al. (2007). The imbalance proportion was set this way because in previous studies, researchers found that models with extreme population mixture proportions (i.e., 10% or 7%) of a subpopulation were less likely to converge and the class enumeration was less accurate (Chen et al., 2007; Henson, et al., 2007; Tofighi & Enders, 2008).

The simulation used a 2 (magnitude of the T_{π} matrix: small or medium; magnitude of $\tau_{\beta 00}$: small or medium) * 2 (number of participants per cluster: 20 or 40 cases) * 2 (number of clusters: 30 or 50 clusters) * 2 (mixing proportions: 50%:50% or 75%:25%) * 2 (ICC: .10 or .20) factorial design to generate the data. 500 replications were generated for each condition using the SAS 9.1 Proc IML procedure, yielding a total of (500 datasets * 32 conditions) 16000 datasets. Each dataset was then analyzed as a two-level model (i.e., false model, ignoring the highest (cluster) level, type = mixture) and a three-level model (i.e., true model, considering the highest (cluster) level, type = twolevel mixture) using Mplus 4.21 Mixture routine (Muthén & Muthén, 2006-2007). The accuracy of classification of individuals, and the accuracy as well as the test of significance (i.e., Type I error rate and statistical power) of the parameter estimates of the model for each subpopulation were evaluated.

Analysis

Similar to in Study One, valid replications (i.e., converged with at least 6% of the observations identified for one of the classes) were selected (it turned out that all 16000 replications were valid), with hit rates and relative biases of parameter estimates under the 32 conditions for both true and false models calculated and examined. Analyses of

variance (ANOVAs) were conducted to determine the contribution of the design factors and all possible interactions.

Results

Results of Study One were summarized in Tables 9-14, and ANOVA results indicating factors with $\eta^2 \geq .01$ were summarized in Table 15.

Hit Rate

Table 9 presents the hit rate of the true and false models under the 32 design conditions. The results show that true model's hit rate ranged from 87% to 95%, whereas false model's hit rate ranged from 79% to 90%. In addition, true model had hit rates of 2% to 6% higher compared to the false model within the same design condition.

For the true model, ANOVA results show that two factors had substantial impact on the hit rate, namely, magnitude of the T_π matrix (F (1, 15968) = 24944.579, p<.001, $\eta^2 = .462$) and mixing proportion (F (1, 15968) = 12506.110, p<.001, $\eta^2 = .232$). As the magnitude of the T_π matrix increased, the hit rate of the true model decreased. On the other hand, as the mixing proportions changed from balanced to unbalanced for the two subpopulations, the hit rate of the true model increased.

For the false model, ANOVA results showed that three factors had substantial impact on the hit rate, namely, mixing proportion (F (1, 15968) = 4466.341, p<.001, $\eta^2 = .157$), magnitude of the T_π matrix (F (1, 15968) = 4028.444, p<.001, $\eta^2 = .141$), and ICC (F (1, 15968) = 3356.696, p<.001, $\eta^2 = .118$). As the magnitude of the T_π matrix increased, the hit rate of the true model decreased. On the other hand, as the mixing proportions changed from balanced to unbalanced for the two subpopulations, the hit rate

of the true model increased. As the ICC magnitude increased, hit rate of the false model decreased.

Two factors, ICC (F (1, 15968) = 3608.102, p<.001, $\eta^2 = .175$) followed by mixing proportion (F (1, 15968) = 631.705, p<.001, $\eta^2 = .031$), had impact on the hit rate difference between the true and false models. As the ICC magnitude increased, the hit rate difference increased. As the mixing proportions changed from balanced to unbalanced, the hit rate difference decreased.

Relative Bias of Fixed Effect Estimates

Table 10 presents the mean relative bias (RB) of fixed effect estimates for true and false models under the 32 design conditions. There were four fixed effect estimates, namely, the mean intercept (I1) and mean slope (S1) for latent class 1 (the slow growing group), and the mean intercept (I2) and mean slope (S2) for latent class 2 (the fast growing group).

For the true model, all mean RBs for all four fixed effect estimates were close to zero. ANOVA results indicated that no η^2 s of the five design factors and their interaction effects were larger than .01 when the RB of I1, I2, S1, and S2 were the dependent variables.

For the false model, all mean RBs for all four fixed effect estimates were within $\pm 5\%$ with four exceptions. The exceptions all occurred for S1 when ICC = 0.2 and cluster size = 20. Three were overestimation of the population value (i.e., 24%, 9%, and 22% respectively), and one was underestimation (i.e., -9%). ANOVA results indicated that no η^2 s of the five design factors and their interaction effects were larger than .01 when the

RB of I1, I2, S1, and S2 were the dependent variables. In other words, none of the design factors had substantial effect on the RB of the fixed effect estimates.

Relative Bias of Level-1 Residual Variance Estimates

Table 11 presents the mean relative bias (RB) of level-1 residual variance estimates σ^2 for true and false models under the 32 design conditions. There were four σ^2 for the repeated measures y.

For both the true and false models, all mean RBs for all four σ^2 were close to zero. ANOVA results indicated that no η^2 s of the five design factors and their interaction effects were larger than .01 when the RB of each σ^2 were the dependent variables.

Relative Bias of Level-2 Variance-Covariance Estimates

Table 12 presents the mean relative bias (RB) of level-1 variance and covariance estimates for true and false models under the 32 design conditions. There were three estimates, namely, the variance of intercept ($\tau_{\pi 00}$), the variance of slope ($\tau_{\pi 11}$), and the covariance between intercept and slope ($\tau_{\pi 10}$).

For the true model, all mean RBs for $\tau_{\pi 00}$, $\tau_{\pi 11}$ and $\tau_{\pi 10}$ were within $\pm 5\%$. ANOVA results indicated that no η^2 s of the five design factors and their interaction effects were larger than .01 when the RB of $\tau_{\pi 00}$, $\tau_{\pi 11}$ and $\tau_{\pi 10}$ were the dependent variables.

For the false model, the overestimation of intercept variance was substantial and there was overestimation for $\tau_{\pi 00}$ under all design conditions. The overestimation ranged from 62% to 278%. ANOVA results showed that there was an interaction effect (F (1, 15968) = 1133.247, p<.001, $\eta^2 = .028$) between two factors, namely, ICC and magnitude

of the T_π matrix. As shown in Figure 1, the overestimation decreased as the magnitude of the T_π matrix increased, and the overestimation was larger when ICC was larger. However, the difference in intercept variance overestimation was larger between different ICC conditions when the magnitude of the T_π matrix was smaller, holding constant all other conditions.

On the other hand, the mean RB was not large for false model $\tau_{\pi 11}$. As shown in Table 12, there were $\tau_{\pi 11}$ underestimation in only three conditions and the values were close to -5%. These underestimation of $\tau_{\pi 11}$ occurred under conditions where cluster number = 30, cluster size = 20, and mixing proportions was balanced. ANOVA results indicated that no η^2 s of the five design factors and their interaction effects were larger than .01 when the RB of $\tau_{\pi 11}$ was the dependent variable.

As shown in Table 12, the mean RB of $\tau_{\pi 10}$ was out of the range of $\pm 5\%$ for 8 conditions but within the range of $\pm 5\%$ for the other 24 conditions. For the 8 conditions, $\tau_{\pi 10}$ was overestimated under 7 conditions. ANOVA results indicated that no η^2 s of the five design factors and their interaction effects were larger than .01 when the RB of $\tau_{\pi 10}$ was the dependent variable.

Relative Bias of Standard Errors of Fixed Effects Estimates

As shown in Table 13, using the cutoff value of $\pm 10\%$ (Hoogland & Boomsma, 1998) for RB of standard errors of fixed effects estimates (i.e., SE_{I1} & SE_{I2} for intercepts, and SE_{S1} & SE_{S2} for slopes), the mean RBs of SE_{I1} and SE_{I2} showed a trend of

underestimation, whereas the mean RBs of SE_{S1} and SE_{S2} showed a trend of overestimation.

ANOVA results showed that RBs of SE_{I1} and SE_{I2} were both affected by three design factors, namely, cluster size (Fs (1, 15968) = 840.561 and 731.261, ps < .001, η^2 s = .046 and .042 respectively), mixing proportions (Fs (1, 15968) = 882.369 and 260.775, ps < .001, η^2 s = .048 and .015 respectively), and magnitude of the T_π matrix (Fs (1, 15968) = 298.624 and 363.441, ps < .001, η^2 s = .016 and .021 respectively). Underestimation of SE_{I1} and SE_{I2} increased as cluster size increased, but decreased as magnitude of the T_π matrix increased. When the mixing proportion of the slow growing class decreased from 50% to 25%, the underestimation of SE_{I1} decreased; whereas the underestimation of SE_{I2} increased as the mixing proportion of the fast growing class increased from 50% to 75%. In other words, the underestimation of standard errors for intercepts is a function of the latent class size; larger class size is associated with larger standard error.The RBs of SE_{S1} and SE_{S2} were both affected by ICC (Fs (1, 15968) = 391.188 and 181.654, ps < .001, η^2 s = .024 and .011 respectively). The overestimation of SE_{S1} and SE_{S2} increased as ICC increased.

Statistical Power to Detect Significant Level-2 Variance and Covariance

Statistical power to detect significant $\tau_{\pi 00}$, $\tau_{\pi 11}$ and $\tau_{\pi 10}$ were examined, since the true parameter values of $\tau_{\pi 00}$, $\tau_{\pi 11}$ and $\tau_{\pi 10}$ were larger than zero. The empirical power was represented by the proportion of significant effects (i.e. Z > 1.96) within each condition. The results are shown in Table 14.

The overall mean power for detecting $\tau_{\pi 00}$, $\tau_{\pi 11}$ and $\tau_{\pi 10}$ was .98 .91 .52 for the true model, and 1.00 .84 .35 for the false model respectively. The POWER to detect $\tau_{\pi 11}$ and $\tau_{\pi 10}$ is higher but equal or lower for detecting $\tau_{\pi 00}$ in true model compared to false model, Because power is a direct function of sample size (i.e., cluster number and cluster size) and effect size (i.e., the magnitude of $\tau_{\pi 00}$, $\tau_{\pi 11}$ and $\tau_{\pi 10}$), it is natural to find for both true and false models, cluster number, cluster size, and the magnitude of the T_{π} matrix had a substantial impact (i.e. main effects or interaction effects) on POWER as shown in Table 14. Besides, the POWER was largest for $\tau_{\pi 00}$, followed by $\tau_{\pi 11}$ and $\tau_{\pi 10}$.

According to ANOVA results, power to detect $\tau_{\pi 00}$, $\tau_{\pi 11}$ and $\tau_{\pi 10}$ under the true model was all affected by cluster number, cluster size, and the magnitude of the T_{π} matrix. However, under the false model, in addition to the previously mentioned factors, the power to detect $\tau_{\pi 10}$ was also affected by ICC (F (1, 15968) = 240.302, $p < .001$, $\eta^2 = .013$). The power decreased as ICC increased. On the other hand, ANOVA results indicated that no η^2 s of the five design factors and their interaction effects were larger than .01 when the power to detect $\tau_{\pi 00}$ was the dependent variables.

Discussion

The purpose of this study was to examine the impact of ignoring a higher level nesting structure in multilevel growth mixture models. The accuracy of classification of individuals, and the accuracy as well as the test of significance (i.e., Type I error rate and statistical power) of the parameter estimates of the model for each subpopulation were examined. Five design factors were considered, namely, magnitude of the T_{π} matrix, number of participants per cluster, number of clusters, mixing proportions, and ICC.

Multilevel data with two subpopulations were generated and then analyzed using the true model (MGMM) and the false model (GMM). The results of Study Two are summarized and explained.

Hit Rate

Ignoring the higher level nesting structure results in less accurate classification of the individuals. It is not surprising that classification becomes less accurate as the variance and covariance of each latent class increases, because the individual growth trajectories overlap more between the two latent classes when the variances of the intercept and slope increase. The result that unbalanced mixing proportion leads to more accurate classification is consistent with the finding by Keng, Leite, and Beretvas (2008). When comparing traditional GMM (i.e. using means of multiple indicators at each time point) and Curve-of-factor Model (i.e. using latent factor with multiple indicators at each time point), Keng et al (2008) found that correct class assignment was higher when proportion in each class was more different.

The magnitude of ICC impacts the classification accuracy of false model and the hit rate difference between true and false models. As the variance ignored at the higher level increases (i.e. ICC increases), the classification accuracy decreases. An important advantage of multilevel modeling is that by modeling the higher level nesting structure (e.g. schools or classrooms), the variation in individual growth trajectories can be decomposed into within- and between-school components (Raudenbush & Bryk, 2002). In the same vein, by partitioning out the variance residing in the higher nesting level, MGMM can estimate the variance at the lower-level (e.g. individuals) more precisely and classify individuals more accurately.

Relative Bias of Fixed Effect Estimates

Although there were a few conditions with large relative bias for S1, ignoring the higher level nesting structure does not affect fixed effects estimates very much. This finding is consistent with findings in previous studies (Moerbeek, 2004; Meyers & Beretvas, 2006) examining the impact of ignoring a higher level nesting in multilevel analysis and ignoring a cross-classified factor in cross-classified data structures.

Relative Bias of Level-1 and Level-2 Variance Estimates

Ignoring the higher level nesting structure affects level-2 variance-covariance estimates, but not level-1 residual variance estimates. This finding is also consistent with findings in previous studies (Moerbeek, 2004).

Because the ignored variance component at level-3 resides in cluster mean intercept (i.e. $\tau_{\beta 00}$), it is natural that this variance is redistributed to level-2 intercept variance, resulting in overestimation of the level-2 intercept variance and little influence on slope variance estimation. In addition, as the variance ignored increased, the overestimation of $\tau_{\pi 00}$ increases. Smaller T_{π} matrix results in larger overestimation, because when the amount of redistributed level-3 variance is the same and all other design conditions are the same, the impact on $\tau_{\pi 00}$ is greater when $\tau_{\pi 00}$ is smaller according to the way RB is calculated.

Relative Bias of Standard Errors of Fixed Effects Estimates

Ignoring the higher level nesting structure affects the standard errors (SE) of intercepts and slopes in different ways. There is a trend of underestimating intercept SEs whereas the slope SEs are generally overestimated. The result related to intercept SE is consistent with the results in Moerbeek's (2004) study. When ignoring the higher level

nesting structure in a model with random intercept and one single population, Moerbeek found that the variance of the intercept was overestimated but the SE of intercept was underestimated. In a study examining the impact of ignoring a cross-classified factor in a model with random intercept and a single population (Meyers and Beretvas, 2006), similar results were also found when a cross-classified factor was ignored. The results of underestimated intercept SE, both from this study and other previous studies is still unclear and further research on this is needed. On the other hand, the slope SE is obviously a function of both slope ($\tau_{\pi 11}$) and intercept ($\tau_{\pi 00}$) variances and the overestimation of the slope SE is probably due to the inflation of $\tau_{\pi 00}$ but not $\tau_{\pi 11}$.

Statistical Power to Detect Significant Level-2 Variance and Covariance

Ignoring the higher level nesting structure results in less power to detect significant intercept-slope covariance $\tau_{\pi 10}$ and slope variance $\tau_{\pi 11}$, but does not influence the power to detect significant intercept variances $\tau_{\pi 00}$. This situation can be explained by the inflation of $\tau_{\pi 00}$ in the false model. Besides, the amount of variance at level-3 influences power.

CHAPTER IV

GENERAL DISCUSSION AND CONCLUSIONS

Summary and Discussion

In this dissertation, the impact of ignoring a higher level nesting structure has been examined in both cross-sectional and longitudinal data. There are three design factors (i.e. R^2, ICC and Imbalance Type) for the cross-sectional study, and five factors (i.e., magnitude of the T_π matrix, number of participants per cluster, number of clusters, mixing proportions, and ICC) for the longitudinal study. The accuracy of classification of individuals, and the accuracy as well as the test of significance (i.e., Type I error rate and statistical power) of the parameter estimates of the model for each subpopulation have been examined. The impact of ignoring a higher level nesting structure in (growth) mixture model is summarized as follows around hit rate, fixed effect estimates, variance estimates, and standard error estimates.

Hit Rate

Classification of individuals is less accurate when ignoring a higher level structure. The accuracy of classification is mostly affected by the difference in fixed effect parameter between the latent classes (i.e., R^2), the amount of variance ignored at the highest level (i.e., ICC), the variance of each latent class (i.e., Imbalance Type, T_π matrix), and the mixing proportion of each latent class (i.e., Imbalance Type, mixing proportion). It is not surprising that factors related to the separation of latent classes (i.e. R^2, Imbalance Type, T_π matrix) affect classification accuracy.

ICC, the factor that measures the proportion of variance in the outcome that is between groups/clusters, has a great impact on classification accuracy. This reinforces the importance of modeling the higher level nesting structure in (growth) mixture modeling. By modeling the higher level nesting structure (e.g. schools or classrooms), the variation in individual growth trajectories can be decomposed into within- and between-cluster components, the variances at the within-cluster level can be better estimated and individuals can be classified more accurately.

It is interesting to find unbalanced mixing proportion actually helps classification in Study Two. This result is a little bit different from that of Simulation 1B in Study One, probably because the unbalanced condition in Simulation 1B is more complicated than the condition in Study Two. In Simulation 1B, the unbalanced latent classes differed not only in mixing proportions, but also in variances. However, the latent classes in Study Two are well-separated classes, whereas in Simulation 1B there were different separation levels. It is likely that the difference in variances of latent classes and the separation level play an important role in classification accuracy in Simulation 1B. Nevertheless, the result in Study Two is consistent with findings by Keng et al. (2008) and this factor is worth more investigation.

Fixed Effect Estimates

Fixed effect estimates are less likely to be affected when a higher level nesting structure is ignored. The accuracy of fixed effect estimates is affected by difference in fixed effect parameter between the latent classes (i.e., R^2) and the amount of variance ignored at the highest level (i.e., ICC). It is not surprising that the factor related to the separation of latent classes (i.e., R^2) affects fixed effect estimates. As shown by Study

One, RBs of fixed effect estimates was less accurate at low levels of R^2 even when the higher level nesting structure is modeled. On the other hand, as shown by results in both studies, the fixed effect estimates are more likely to be biased under conditions with higher levels of ICC when all other conditions stay the same.

Variance Estimates

Ignoring a higher level structure will cause the variance at the highest level (i.e., school) to be redistributed to next adjacent level (i.e. student), but not the lowest level (i.e. repeated measures), as shown in Study Two. The variance estimation bias is affected by ICC and factors related to variance of each latent class (i.e., Imbalance Type and T_π matrix), as according to the RB calculation method, the RB is a function of the numerator (i.e., the difference between variance at the highest level and the level-1 variance) and denominator (level-1 variance). In addition, the statistical power to detect significant covariance between the intercept and slope decreased as ICC increased according to the finding in Study Two.

Standard Error Estimates

In Study One, there is a trend of inflation for SE of group mean estimates; whereas in Study Two, there is a trend of underestimation for SE for intercept estimates and inflation for SE of slope estimates. The major difference between the two studies on the random effect design is that in Study One, there is only one random effect associated with each level; whereas in Study Two, there are two random effects at level-2 and there are three variance-covariance parameters. The seemingly contradictory results may be due to the complexity of the variance-covariance matrix of the random effects in Study Two and

further research on this is needed. As a result, the statistical power to detect the significant fixed effect is compromised.

Implications

These findings have practical implications for researchers. According to the findings of the study, ignoring a higher level nesting structure results in lower classification accuracy, less accurate fixed effect estimates, inflation of lower-level variance estimates, and less accurate SE estimates. Therefore, it is important to model the higher level nesting structure in (growth) mixture modeling. Since ICC is an important influencing factor when ignoring the higher level nesting structure, it would be good for researchers to calculate or estimate the magnitude of ICC before conducting the analysis. If the ICC is larger than 0.1, it probably is not suitable to ignore the higher level nesting structure.

In real data analysis, researchers seldom know in advance the true parameter values (i.e. the difference between groups, the true variance and mixing proportions of each latent class) and the class membership. Therefore, in order to have more accurate classification and parameter estimates, and the correct statistical test to detect significant effect, it is important to model the nesting structure and use multilevel (growth) mixture model.

In addition, there are situations when it is difficult to take into account the nesting structure, such as the lack of identifiers on all possible levels of nesting in data (Moerbeek, 2004) and the difficulty in achieving convergence in model estimation (Van Landeghem, Fraine & Damme, 2005). If this is the case, researchers should then be cautious in interpret the findings, especially when they have a marginally significant test result, because it might be a significant result if the researchers considered the nesting structure of the data,

or it might just be a spurious significant results due to the underestimation of standard error.

Limitations and Suggestions for Future Research

The first limitation of the dissertation is that a major assumption made in both studies is that both the true and false models can actually detect there are two latent classes so that the focus is to compare what happens after two latent classes are uncovered. However, it is possible that the false model might not be able to recover the correct number of classes.

Second, only strictly hierarchical data structure is examined in both studies, whereas in reality some data structure is not strictly hierarchical. They are cross-classified in the sense that students come from varied combinations of higher level nesting factors such as schools and neighborhoods. Researchers have found that ignoring the cross-classified structure will result in bias in standard error estimates although the fixed effects estimates were not affected (Van Landeghem, Fraine & Damme, 2005; Meyers & Beretvas, 2006; Luo & Kwok, 2006). However, there is no software available in the area of structural equation modeling to take into account the cross-classified structure in multilevel (growth) mixture modeling. More research and advances in software is needed for the area of multilevel (growth) mixture modeling.

REFERENCES

Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling* (pp. 243–277). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Arminger, G. & Stein, P. (1997). Finite mixtures of covariance structure models with regressors. *Sociological Methods & Research, 26*, 148-182.

Asparouhov, T. & Muthen, B. (2008). Multilevel mixture models. In G.R. Hancock, & K.M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 27-51). Greenwich, CT: Information Age.

Bauer, D. J., & Curran, P. J. (2003a). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods, 8*(3), 338-363.

Bauer, D. J., & Curran, P. J. (2003b). Overextraction of latent trajectory classes: Much ado about nothing? Reply to Rindskopf (2003), Muthén (2003), and Cudeck and Henly (2003). *Psychological Methods, 8*(3) 384-393.

Bauer, D. J., & Curran, P. J. (2004). The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. *Psychological Methods, 9*(1) 3-29.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. Hoboken, NJ: John Wiley & Sons.

Boscardin, C. K., Muthén, B., Francis, D. J., & Baker, E. L. (2008). Early identification of reading difficulties using heterogeneous developmental trajectories. *Journal of Educational Psychology, 100*(1), 192-208.

Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification, 13*, 195-212.

Chen, Q., Kwok, O. & Luo, W. (2007, April). *The impact of ignoring a level of nesting structure in multilevel mixture models: A Monte Carlo study*. Paper presented at the annual meeting of American Educational Research Association, Chicago.

Chen, Q., & Willson, V. L. (2006, June). *Effect of covariate magnitude on latent class analysis determination*. Paper presented at the 71st Annual Meeting of the Psychometric Society (IMPS2006), Montréal, Canada.

Clogg, C. C. (1995). Latent class models. In G. Arminger, C. C. Clogg, and M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 311-359). New York: Plenum.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd Ed.). Mahwah, NJ: Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.

Collins, L. M., & Flaherty, B. P. (2002). Latent class models for longitudinal data. In. J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 287-303). Cambridge: Cambridge University Press.

Connell, A. M., & Frye, A. A. (2006). Growth mixture modeling in developmental psychology: Overview and demonstration of heterogeneity in developmental trajectories of adolescent antisocial behavior. *Infant and Child Development, 15*(6), 609-621.

Courvoisier, D. S., Eid, M., & Nussbeck, F. W. (2007). Mixture distribution latent state-trait analysis: Basic ideas and applications. *Psychological Methods, 12*(1), 80-104.

D'Angiulli, A., Siegel, L. S., & Maggi, S. (2004). Literacy instruction, SES, and word-reading achievement in English-language learners and children with English as a first language: A longitudinal study. *Learning Disabilities Research & Practice, 19,* 202-213.

Dempster, A. P., Laird, M. N. & Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the E-M algorithm. *Journal of the Royal Statistical Society, Series B, 39*, 1-38.

Dolan, C. V., & van der Maas, H. L. J. (1998). Fitting multivariate normal finite mixtures subject to structural equation modeling. *Psychometrika, 63*, 227–253.

Duncan, T. E., Duncan, S. C., Strycker, L. A., Li, F., & Alpert, A. (1999). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications*. Mahwah, NJ: Erlbaum.

Enders, C. K., & Tofighi, D. (2008). The impact of misspecifying class-specific residual variances in growth mixture models. *Structural Equation Modeling, 15*(1), 75-95.

Everitt, B. S. & Hand, D. J. (1981).*Finite mixture distributions*. New York: Chapman and Hall Ltd.

Everitt, B. S., Landau, S. & Leese, M. (2001). *Cluster analysis*. New York: Oxford University Press.

Ferron, J., Dailey, R., & Yi, Q. (2002). Effects of misspecifying the first-level error structure in two-level models of change. *Multivariate Behavioral Research, 37*, 379-403.

Henson, J. M., Reise, S. P., & Kim, K. H. (2007). Detecting mixtures from structural model differences using latent variable mixture modeling: A comparison of relative model fit statistics. *Structural Equation Modeling, 14*(2), 202-226.

Hipp, J. R., & Bauer, D. J. (2006). Local solutions in the estimation of growth mixture models. *Psychological Methods, 11*(1), 36-53.

Hoogland, J.J. & Boomsma, A. (1998). Robustness studies in covariance structure modeling. *Sociological Methods and Research, 26*, 329-367.

Hox, J. (2002). *Multilevel analysis techniques and applications*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers.

Jedidi, K., Jagpal, H. S., & DeSarbo, W. S. (1997). Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science, 16*, 39-59.

Jones, B. L., Nagin, D. S., & Roeder, K. (2001). A SAS procedure based on mixture models for estimating developmental trajectories. *Sociological Methods &Research, 29*, 374–393.

Keng, L, Leite, W. L. & Beretvas , S. N (2008, March). *Comparing growth mixture models when measuring latent constructs with multiple indicators*. Paper presented at the annual meeting of American Educational Research Association, New York.

Keselman, H. J., Algina, J., Kowalchuk, R. K., & Wolfinger, R. D. (1998). A comparison of two approaches for selecting covariance structures in the analysis of repeated measurements. *Communications in Statistics-Simulation and Computation, 27*, 591-604.

Khoo, S., West, S. G., Wu, W., & Kwok, O. (2006). Longitudinal methods. In M. Eid & E. Diener (Eds.), *Handbook of psychological measurement: A multimethod perspective* (pp.301-317). Washington, DC: APA.

Lazarsfeld, P., & Henry, N. (1968). *Latent structure analysis*. New York: Houghton Mifflin.

Luo, W. & Kwok, O. (2006, June). *Impacts of ignoring a crossed factor in analyzing multilevel data with cross-classified structures*. Paper presented at the annual meeting of Psychometric Society, Montréal, Canada.

McLachlan, G. (1992). *Discriminant analysis and statistical pattern recognition*. New York: Wiley.

McLachlan, G., & Basford, K. (1988). *Mixture models*. New York: Marcel Dekker, Inc.

McLachlan, G., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika, 55*, 107-122.

Meyers, J. L., & Beretvas, S. N. (2006). The impact of inappropriate modeling of cross-classified data structures. *Multivariate Behavioral Research, 41*, 473-497.

Moerbeek, M. (2004). The consequence of ignoring a level of nesting in multilevel analysis. *Multivariate Behavioral Research, 39*, 129-149.

Muthén, B. (2001a). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class/latent growth modeling. In L.M. Collins & A. Sayer (Eds.), *New methods for the analysis of change* (pp. 291-322). Washington, D.C.: APA.

Muthén, B. (2001b). Latent variable mixture modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 1-33). Mahwah, NJ: Lawrence Erlbaum Associates.

Muthén, B. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika, 29*, 81-117.

Muthén, B. (2003). Statistical and substantive checking in growth mixture modeling: Comment on Bauer and Curran (2003). *Psychological Methods, 8*(3), 369-377.

Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 345-368). Newbury Park, CA: Sage Publications.

Muthén, B. (2006a). The potential of growth mixture modeling. *Infant and Child Development, 15*(6), 623-625.

Muthen, B. (2006b, August). New longitudinal analysis techniques using Mplus. Workshop given at UCLA Summer Institute on Longitudinal Research, Los Angeles, CA.

Muthén, B. (2008). Latent variable hybrids: Overview of old and new models. In G. R. Hancock & K. M. Samuelsen (Eds.), (2008). *Advances in latent variable mixture models*. Charlotte, NC: Information Age Publishing, Inc.

Muthén, B. O. (2005, September 21). *Multilevel mixture model*. Message posted to http://www.statmodel.com/discussion/messages/13/809.html?1143606816

Muthén, B. O., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research, 24*(6), 882-891.

Muthen, B. O. & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics, 55*, 463-469.

Muthén, L. K. (2003, March 28). *Multilevel mixture model*. Message posted to http://www.statmodel.com/discussion/messages/14/268.html?1069377168

Muthén, L. K., & Muthén, B. O. (2006-2007). *Mplus user's guide (V4.21)*. Los Angeles, CA: Muthén & Muthén.

Nagin, D. S. (1999). Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological Methods, 4*, 139-157.

Nagin, D. S. (2005). *Group-based modeling of development*. Cambridge, MA: Harvard University Press.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14*(4), 535-569.

Palardy, G., & Vermunt, J. K. (2007). *Multilevel growth mixture models for classifying group-level observations*. Retrieved July 7, 2007, from http://spitswww.uvt.nl/~vermunt/

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods.* Thousand Oaks, CA: Sage.

Raudenbush, S. W., & Liu, X. (2001). Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods, 6,* 387-401.

SAS Institute Inc. (2002-2004). *SAS, release 9.1.3*. [Computer program] Cary, NC: Author.

Snijders, T. & Bosker, R. (1999). *Multilevel analysis*. Thousand Oaks, CA: Sage.

Tofighi, D., & Enders, C.K. (2008). Identifying the correct number of classes in a growth mixture models. In G.R. Hancock, & K.M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 317-341). Greenwich, CT: Information Age.

Van Landeghem, G., De Fraine, B., & Van Damme, J. (2005). The consequence of ignoring a level of nesting in multilevel analysis: A comment. *Multivariate Behavioral Research, 40*, 423-434.

Vermunt, J.K. (2007). Latent class and finite mixture models for multilevel data sets. *Statistical Methods in Medical Research*. Retrieved July 7, 2007, from http://spitswww.uvt.nl/~vermunt/

Vermunt, J.K., & Magidson, J. (2002). Latent class cluster analysis. In J. Hagenaars & A. McCutcheon (Eds.), *Applied latent class analysis* (pp. 89-106). New York: Cambridge University Press.

Vermunt, J.K. and Magidson, J. (2004a). Latent class analysis. In M. Lewis-Beck, A. Bryman, and T.F. Liao (Eds.), *The Sage encyclopedia of social sciences research methods* (pp. 549-553). Thousand Oakes, CA: Sage.

Vermunt, J.K. and Magidson, J. (2004b). Latent variable. In M.S. Lewis-Beck, A. Bryman, and T.F. Liao (Eds.), *The Sage encyclopedia of social sciences research methods* (pp.555-556). Thousand Oakes, CA: Sage.

Vermunt, J.K. and Magidson, J. (2004c). Local independence. M. Lewis-Beck, A. Bryman, and T.F. Liao (Eds.), M.S. Lewis-Beck, A. Bryman, and T.F. Liao (Eds.), *The Sage Encyclopedia of Social Sciences Research Methods* (pp.580-581). Thousand Oakes, CA: Sage Publications.

Vermunt, J.K. and Magidson, J. (2005). Structural equation models: Mixture models. In B. Everitt and D. Howell, (Eds.), *Encyclopedia of statistics in behavioral science* (pp.1922–1927). Chichester, UK: Wiley.

Wampold, B. E., & Serlin, R. C. (2000). The consequence of ignoring a nested factor on measures of effect size in analysis of variance. *Psychological Methods, 5*(4), 425-433.

West, S. G., Biesanz, J. C., & Kwok, O. (2003). Within subject and longitudinal experiments: design and analysis issues. In C. Sansone, C. C. Morf, & A. T. Panter (Eds.), *Handbook of methods in social psychology* (pp. 287-312). Thousand Oaks, CA: Sage.

APPENDIX A

TABLES

Table1

Hit Rate of True and False Models in Simulation 1A

Conditions		Valid Replications	Hit Rate		
R^2	ICC		True	False	Differ
0.1	0.1	258	61%	61%	1%
0.1	0.2	217	61%	60%	2%
0.3	0.1	411	73%	70%	3%
0.3	0.2	390	73%	68%	4%
0.5	0.1	496	84%	81%	3%
0.5	0.2	491	84%	78%	6%

Note. Differ = True Model Hit Rate – False Model Hit Rate.

Table 2

Relative Bias of Group Mean Estimates in Simulation 1A

Conditions		True		False	
R^2	ICC	Class1	Class2	Class1	Class2
0.1	0.1	-23%	15%	-24%	17%
0.1	0.2	-23%	16%	-28%	20%
0.3	0.1	-5%	3%	-5%	3%
0.3	0.2	-6%	3%	-9%	4%
0.5	0.1	-2%	1%	0%	0%
0.5	0.2	-3%	1%	-1%	0%

Table 3

Relative Bias of Variance Estimates in Simulation 1A

Conditions		True		False
R^2	ICC	σ^2	τ_{00}	σ^2
0.1	0.1	-20%	-9%	-11%
0.1	0.2	-20%	-5%	0%
0.3	0.1	-5%	-16%	11%
0.3	0.2	-6%	-7%	26%
0.5	0.1	-1%	-16%	27%
0.5	0.2	-3%	-9%	53%

Table 4

Relative Bias of Standard Errors of Group Mean Estimates in Simulation 1A

Conditions		False Model S.E. Bias	
R^2	ICC	SE1	SE2
0.1	0.1	9%	17%
0.1	0.2	11%	8%
0.3	0.1	20%	31%
0.3	0.2	21%	20%
0.5	0.1	3%	9%
0.5	0.2	13%	13%

Table 5
Hit Rate of True and False Models in Simulation 1B

Conditions			Valid	Average Hit Rate		
Imbalance	R^2	ICC	Replications	True	False	Differ
1	0.1	0.1	134	52%	51%	1%
1	0.1	0.2	98	53%	49%	3%
1	0.3	0.1	320	72%	67%	6%
1	0.3	0.2	276	72%	63%	10%
1	0.5	0.1	431	87%	82%	6%
1	0.5	0.2	341	87%	76%	11%
2	0.1	0.1	176	77%	76%	1%
2	0.1	0.2	146	77%	75%	3%
2	0.3	0.1	401	83%	81%	1%
2	0.3	0.2	356	83%	79%	3%
2	0.5	0.1	496	87%	85%	2%
2	0.5	0.2	479	88%	83%	4%

Note.
Imbalance Type 1: Class 1—large size large variance, Class 2—small size small variance
Imbalance Type 2: Class 1—large size small variance, Class 2—small size large variance

Table 6
Relative Bias of Group Mean Estimates in Simulation 1B

Conditions			True Model Class Mean Bias		False Model Class Mean Bias	
Imbalance	R^2	ICC	Class 1	Class 2	Class 1	Class 2
1	0.1	0.1	-64%	-5%	-66%	-5%
1	0.1	0.2	-64%	-3%	-69%	-2%
1	0.3	0.1	-23%	-3%	-33%	-7%
1	0.3	0.2	-24%	-3%	-44%	-8%
1	0.5	0.1	-2%	1%	-6%	-1%
1	0.5	0.2	-3%	1%	-14%	-3%
2	0.1	0.1	-3%	0%	-3%	0%
2	0.1	0.2	-3%	2%	-4%	1%
2	0.3	0.1	1%	15%	1%	14%
2	0.3	0.2	-1%	14%	-3%	10%
2	0.5	0.1	0%	4%	1%	6%
2	0.5	0.2	0%	5%	1%	6%

Note.
Imbalance Type 1: Class 1—large size large variance, Class 2—small size small variance
Imbalance Type 2: Class 1—large size small variance, Class 2—small size large variance

Table 7
Relative Bias of Variance Estimates in Simulation 1B

Conditions			True			False	
Imbalance	R^2	ICC	σ_1^2	σ_2^2	τ_{00}	σ_1^2	σ_2^2
1	0.1	0.1	-26%	50%	-7%	-16%	78%
1	0.1	0.2	-29%	47%	-7%	-7%	110%
1	0.3	0.1	-16%	18%	-12%	-7%	71%
1	0.3	0.2	-16%	14%	-8%	2%	117%
1	0.5	0.1	-2%	-4%	-11%	17%	67%
1	0.5	0.2	-5%	-3%	-11%	28%	133%
2	0.1	0.1	-6%	-13%	-11%	-1%	2%
2	0.1	0.2	-7%	-13%	-10%	2%	17%
2	0.3	0.1	-17%	0%	-8%	-9%	24%
2	0.3	0.2	-18%	-1%	-5%	1%	43%
2	0.5	0.1	-8%	-1%	-7%	0%	37%
2	0.5	0.2	-9%	0%	-5%	11%	74%

Note.
Imbalance Type 1: Class 1—large size large variance, Class 2—small size small variance
Imbalance Type 2: Class 1—large size small variance, Class 2—small size large variance

Table 8
Relative Bias of Standard Errors of Group Mean Estimates in Simulation 1B

Conditions			False Model SE Bias	
Imbalance	R^2	ICC	SE1	SE2
1	0.1	0.1	-21%	17%
1	0.1	0.2	14%	33%
1	0.3	0.1	14%	20%
1	0.3	0.2	32%	32%
1	0.5	0.1	47%	63%
1	0.5	0.2	75%	90%
2	0.1	0.1	2%	-5%
2	0.1	0.2	4%	-18%
2	0.3	0.1	-13%	-19%
2	0.3	0.2	0%	-8%
2	0.5	0.1	-21%	-1%
2	0.5	0.2	-16%	16%

Note.
Imbalance Type 1: Class 1—large size large variance, Class 2—small size small variance
Imbalance Type 2: Class 1—large size small variance, Class 2—small size large variance

Table 9
Mean Hit Rate for True and False Models

Conditions					Hit Rate		
Mixing%	ICC	τ_π	Cluster#	Clus Size	True	False	Differ
50%: 50%	0.1	Small	30	20	90%	87%	3%
				40	90%	88%	3%
			50	20	90%	88%	3%
				40	91%	88%	3%
		Medium	30	20	87%	84%	3%
				40	87%	85%	3%
			50	20	87%	84%	3%
				40	87%	85%	3%
	0.2	Small	30	20	90%	84%	6%
				40	91%	85%	6%
			50	20	90%	85%	6%
				40	91%	85%	5%
		Medium	30	20	87%	79%	8%
				40	87%	81%	6%
			50	20	87%	81%	6%
				40	88%	82%	5%
25%: 75%	0.1	Small	30	20	92%	90%	2%
				40	92%	90%	2%
			50	20	92%	90%	2%
				40	92%	90%	2%
		Medium	30	20	89%	87%	2%
				40	90%	88%	2%
			50	20	90%	88%	2%
				40	90%	88%	2%
	0.2	Small	30	20	92%	87%	5%
				40	92%	88%	4%
			50	20	92%	88%	4%
				40	92%	88%	4%
		Medium	30	20	89%	84%	5%
				40	90%	85%	4%
			50	20	90%	85%	4%
				40	90%	86%	4%

Note. Differ=True Model Hit Rate – False Model Hit Rate.

Table 10
Mean Relative Bias of Fixed Effect Estimates for True and False Models

Conditions					True Model Fixed Effect Bias				False Model Fixed Effect Bias			
Mixing%	ICC	τ_π	Cluster#	ClusSize	I1	I2	S1	S2	I1	I2	S1	S2
50%: 50%	0.1	Small	30	20	-2%	0%	1%	0%	0%	0%	2%	0%
				40	-1%	0%	1%	0%	0%	0%	2%	0%
			50	20	-1%	1%	-3%	0%	0%	0%	-2%	0%
				40	-1%	0%	-1%	0%	0%	0%	-1%	0%
		Medium	30	20	-1%	1%	2%	0%	1%	0%	1%	-2%
				40	-1%	0%	-1%	0%	-1%	0%	-1%	0%
			50	20	-1%	1%	-1%	0%	0%	0%	0%	0%
				40	-1%	0%	-1%	0%	0%	0%	0%	-1%
	0.2	Small	30	20	-2%	0%	-1%	0%	0%	-1%	-3%	0%
				40	0%	1%	-1%	0%	0%	0%	-2%	0%
			50	20	-1%	1%	-1%	0%	1%	0%	3%	0%
				40	-1%	0%	-1%	0%	-1%	0%	0%	0%
		Medium	30	20	-2%	1%	-1%	-1%	2%	-1%	24%	-4%
				40	0%	1%	1%	0%	0%	0%	4%	-2%
			50	20	-1%	1%	-3%	1%	1%	0%	9%	0%
				40	-2%	0%	0%	0%	0%	0%	4%	0%
25%: 75%	0.1	Small	30	20	-2%	0%	0%	0%	0%	0%	0%	0%
				40	-1%	0%	0%	0%	0%	0%	0%	0%
			50	20	-2%	0%	-1%	0%	0%	0%	-1%	0%
				40	-1%	0%	0%	0%	0%	0%	1%	0%
		Medium	30	20	-3%	1%	0%	0%	-1%	0%	2%	0%
				40	-2%	0%	2%	0%	-1%	0%	-1%	0%
			50	20	-4%	0%	-1%	0%	-2%	0%	-3%	0%
				40	-1%	0%	0%	0%	0%	0%	0%	0%
	0.2	Small	30	20	-3%	0%	1%	0%	-1%	0%	1%	0%
				40	-1%	0%	-3%	0%	0%	0%	-2%	0%
			50	20	-2%	0%	-3%	0%	0%	0%	-9%	0%
				40	-1%	0%	-2%	0%	0%	0%	-2%	0%
		Medium	30	20	-3%	0%	4%	0%	0%	0%	22%	-1%
				40	-2%	0%	-4%	0%	-1%	0%	-5%	0%
			50	20	-3%	1%	-2%	0%	0%	0%	-3%	0%
				40	-2%	0%	0%	0%	0%	0%	4%	0%

Note. I = Intercept, S = Slope;
1 = Latent Class 1 (Slow Growing Class), 2 = Latent Class 2 (Fast Growing Class).

Table 11
Mean Relative Bias of Level-1 Residual Variance Estimates for True and False Models

Conditions					True Model Level-1 Residual Variance Bias				False Model Level-1 Residual Variance Bias			
Mixing%	ICC	τ_π	Cluster#	Clus Size	σ_1^2	σ_2^2	σ_3^2	σ_4^2	σ_1^2	σ_2^2	σ_3^2	σ_4^2
50%: 50%	0.1	Small	30	20	0%	0%	0%	1%	0%	0%	0%	1%
				40	0%	0%	0%	0%	0%	0%	0%	0%
			50	20	0%	0%	0%	0%	0%	0%	0%	0%
				40	0%	0%	0%	0%	0%	0%	0%	0%
		Medium	30	20	1%	0%	0%	1%	1%	0%	0%	0%
				40	0%	0%	0%	0%	0%	0%	0%	0%
			50	20	0%	0%	0%	0%	0%	0%	0%	0%
				40	0%	0%	0%	0%	0%	0%	0%	0%
	0.2	Small	30	20	-1%	0%	0%	0%	-1%	0%	0%	0%
				40	0%	0%	0%	0%	0%	0%	0%	0%
			50	20	0%	0%	0%	0%	0%	0%	-1%	0%
				40	0%	0%	0%	0%	0%	0%	0%	0%
		Medium	30	20	0%	0%	0%	0%	0%	0%	0%	0%
				40	0%	0%	0%	1%	0%	0%	0%	1%
			50	20	0%	0%	0%	0%	0%	0%	0%	0%
				40	0%	0%	0%	0%	0%	0%	0%	0%
25%: 75%	0.1	Small	30	20	0%	0%	0%	0%	0%	0%	0%	0%
				40	0%	0%	0%	0%	0%	0%	0%	0%
			50	20	0%	0%	0%	0%	0%	0%	0%	0%
				40	0%	0%	0%	0%	0%	0%	0%	0%
		Medium	30	20	0%	0%	0%	1%	0%	0%	0%	1%
				40	0%	0%	0%	0%	0%	0%	0%	0%
			50	20	-1%	0%	0%	0%	-1%	0%	0%	0%
				40	0%	0%	0%	0%	0%	0%	0%	0%
	0.2	Small	30	20	-1%	0%	0%	0%	0%	0%	0%	0%
				40	0%	0%	0%	0%	0%	0%	0%	1%
			50	20	0%	0%	0%	0%	0%	0%	0%	0%
				40	0%	0%	0%	0%	0%	0%	0%	0%
		Medium	30	20	0%	0%	0%	0%	0%	0%	1%	0%
				40	0%	0%	0%	0%	0%	0%	0%	0%
			50	20	0%	0%	0%	0%	0%	0%	0%	0%
				40	0%	0%	0%	0%	0%	0%	0%	0%

Note. σ_1^2 refers to the residual variance of y at Time 1;
Subscripts 1, 2, 3, and 4 refer to Time 1, 2, 3, and 4 respectively.

Table 12
Mean Relative Bias of Level-2 Variance and Covariance Estimates for True and False Models

Conditions					True Model Level-2 Variance Bias			False Model Level-2 Variance Bias		
Mixing%	ICC	τ_{π}	Cluster#	ClusSize	$\tau_{\pi 10}$	$\tau_{\pi 00}$	$\tau_{\pi 11}$	$\tau_{\pi 10}$	$\tau_{\pi 00}$	$\tau_{\pi 11}$
50%: 50%	0.1	Small	30	20	-5%	-2%	-4%	-1%	122%	-7%
				40	-1%	-2%	-1%	3%	117%	0%
			50	20	-3%	-3%	-2%	-3%	115%	-2%
				40	3%	-2%	1%	4%	120%	1%
		Medium	30	20	2%	-3%	-3%	15%	66%	-3%
				40	-2%	-2%	-1%	-1%	62%	-1%
			50	20	-3%	-1%	-2%	0%	67%	-2%
				40	1%	-1%	1%	3%	64%	1%
	0.2	Small	30	20	-4%	-5%	-3%	10%	278%	-6%
				40	2%	-3%	-1%	3%	260%	-2%
			50	20	-4%	-5%	-2%	8%	273%	-3%
				40	-3%	-3%	0%	2%	268%	1%
		Medium	30	20	0%	-4%	-3%	28%	158%	-6%
				40	0%	-2%	-2%	15%	148%	-2%
			50	20	-3%	-3%	-2%	16%	153%	-2%
				40	-2%	-2%	0%	5%	147%	0%
25%: 75%	0.1	Small	30	20	0%	-4%	-4%	2%	118%	-4%
				40	-4%	-3%	-5%	-5%	117%	-5%
			50	20	2%	-3%	-2%	2%	118%	-3%
				40	1%	-2%	-2%	1%	121%	-2%
		Medium	30	20	0%	-2%	-2%	2%	64%	-3%
				40	0%	-3%	-1%	1%	63%	-2%
			50	20	0%	-2%	-1%	0%	65%	-1%
				40	-2%	-1%	0%	-2%	66%	-1%
	0.2	Small	30	20	-2%	-3%	-4%	0%	263%	-5%
				40	-4%	-2%	-3%	-4%	258%	-4%
			50	20	-4%	-4%	-2%	-8%	267%	-5%
				40	-1%	-2%	-1%	-2%	265%	-2%
		Medium	30	20	2%	-1%	-4%	18%	146%	-5%
				40	-2%	-1%	-1%	1%	142%	-2%
			50	20	-2%	-3%	-1%	4%	149%	-1%
				40	-1%	-2%	0%	3%	146%	0%

Note. $\tau_{\pi 10}$ = Intercept Slope Covariance, $\tau_{\pi 00}$ = Intercept Variance, $\tau_{\pi 11}$ = Slope Variance.

Table 13
Mean Relative Bias of Standard Errors of Fixed Effect Estimates for False Models

Conditions					False Model S.E. Bias			
Mixing%	ICC	τ_π	Cluster#	Clus Size	SE_{I1}	SE_{I2}	SE_{S1}	SE_{S2}
50%: 50%	0.1	Small	30	20	-15%	-15%	14%	18%
				40	-37%	-37%	11%	16%
			50	20	-18%	-20%	14%	15%
				40	-40%	-38%	19%	26%
		Medium	30	20	3%	-6%	22%	26%
				40	-26%	-28%	21%	26%
			50	20	-8%	-10%	25%	24%
				40	-30%	-27%	13%	21%
	0.2	Small	30	20	-11%	-16%	54%	48%
				40	-40%	-40%	43%	37%
			50	20	-20%	-19%	48%	38%
				40	-42%	-42%	38%	37%
		Medium	30	20	-7%	7%	23%	30%
				40	-17%	-20%	71%	70%
			50	20	0%	1%	51%	41%
				40	-30%	-27%	48%	48%
25%: 75%	0.1	Small	30	20	-1%	-34%	16%	12%
				40	-11%	-47%	17%	12%
			50	20	1%	-30%	12%	15%
				40	-16%	-47%	16%	10%
		Medium	30	20	11%	-15%	18%	12%
				40	-2%	-34%	22%	14%
			50	20	6%	-20%	13%	16%
				40	-7%	-35%	22%	17%
	0.2	Small	30	20	10%	-31%	29%	37%
				40	-19%	-53%	41%	30%
			50	20	17%	-29%	37%	39%
				40	-19%	-53%	41%	25%
		Medium	30	20	23%	-19%	9%	8%
				40	-2%	-38%	43%	38%
			50	20	32%	-14%	46%	44%
				40	4%	-35%	37%	33%

Note. I = Intercept, S = Slope;
1 = Latent Class 1 (Slow Growing Class), 2 = Latent Class 2 (Fast Growing Class).

Table 14
Mean Power to Detect Significant Level-2 Variance and Covariance for True and False Models

Conditions					True Model Power			False Model Power		
Mixing%	ICC	τ_{π}	Cluster#	Clus Size	$\tau_{\pi 10}$	$\tau_{\pi 00}$	$\tau_{\pi 11}$	$\tau_{\pi 10}$	$\tau_{\pi 00}$	$\tau_{\pi 11}$
50%: 50%	0.1	Small	30	20	20%	86%	55%	12%	100%	46%
				40	36%	100%	89%	25%	100%	81%
			50	20	27%	99%	80%	19%	100%	73%
				40	61%	100%	99%	43%	100%	97%
		Medium	30	20	35%	100%	94%	25%	99%	85%
				40	65%	100%	100%	49%	100%	98%
			50	20	54%	100%	99%	39%	100%	96%
				40	90%	100%	100%	77%	100%	100%
	0.2	Small	30	20	18%	88%	60%	8%	100%	44%
				40	38%	100%	88%	17%	100%	71%
			50	20	28%	98%	80%	14%	100%	61%
				40	53%	100%	98%	27%	100%	94%
		Medium	30	20	34%	100%	93%	21%	100%	76%
				40	69%	100%	100%	37%	100%	93%
			50	20	56%	100%	99%	30%	100%	93%
				40	88%	100%	100%	57%	100%	99%
25%: 75%	0.1	Small	30	20	21%	90%	64%	14%	100%	55%
				40	41%	100%	91%	29%	100%	86%
			50	20	36%	100%	83%	26%	100%	78%
				40	63%	100%	99%	47%	100%	98%
		Medium	30	20	44%	100%	96%	33%	100%	89%
				40	77%	100%	100%	62%	100%	100%
			50	20	66%	100%	100%	55%	100%	98%
				40	92%	100%	100%	83%	100%	100%
	0.2	Small	30	20	24%	90%	62%	10%	99%	48%
				40	41%	100%	92%	17%	100%	79%
			50	20	33%	99%	85%	15%	100%	69%
				40	63%	100%	100%	33%	100%	96%
		Medium	30	20	46%	100%	95%	29%	99%	86%
				40	75%	100%	100%	45%	100%	99%
			50	20	65%	100%	100%	38%	100%	97%
				40	92%	100%	100%	71%	100%	100%

Note $\tau_{\pi 10}$ = Intercept Slope Covariance, $\tau_{\pi 00}$ = Intercept Variance, $\tau_{\pi 11}$ = Slope Variance.

Table 15

$\eta^2 \geq .01$ *Values from 5-Way ANOVA for True and False Models*

	Model	Parameter	Mixing%	ICC	τ_π	Cluster#	Clus Size	Interaction
Hit Rate	True		0.232		0.462			
	False		0.157	0.118	0.141			
	Difference		0.031	0.175				
Fixed Effects Relative Bias	True	I1/I2/ S1/S2	No $\eta^2 \geq .01$					
	False	I1/I2/ S1/S2	No $\eta^2 \geq .01$					
Level-1 Residual Variance Relative Bias	True	$\sigma_1^2 / \sigma_2^2 /$ σ_3^2 / σ_4^2	No $\eta^2 \geq .01$					
	False	$\sigma_1^2 / \sigma_2^2 /$ σ_3^2 / σ_4^2	No $\eta^2 \geq .01$					
Level-2 Variance-Covariance Estimates Relative Bias	True	$\tau_{\pi 10}$ $\tau_{\pi 00}$ $\tau_{\pi 11}$	No $\eta^2 \geq .01$					
	False	$\tau_{\pi 10}$	No $\eta^2 \geq .01$					
		$\tau_{\pi 00}$		0.370	0.202			ICC* τ_π (.028)
		$\tau_{\pi 11}$	No $\eta^2 \geq .01$					
False Model SE Bias		SE_{I1}	0.048		0.016		0.046	
		SE_{I2}	0.015		0.021		0.042	
		SE_{S1}		0.024				
		SE_{S2}		0.011				
Power to Detect Significant Variance-Covariance	True	$\tau_{\pi 10}$			0.077	0.032	0.075	#*size(.011)
		$\tau_{\pi 00}$			0.015	0.012	0.016	#* τ_π (.011) size* τ_π (.015)
		$\tau_{\pi 11}$			0.073	0.024	0.051	#* τ_π (.012) size* τ_π (.031)
	False	$\tau_{\pi 10}$		0.013	0.067	0.026	0.047	
		$\tau_{\pi 00}$	No $\eta^2 \geq .01$					
		$\tau_{\pi 11}$			0.081	0.032	0.064	size* τ_π (.018)

Note. Naming conventions follow those from Table 9-15

APPENDIX B

FIGURES

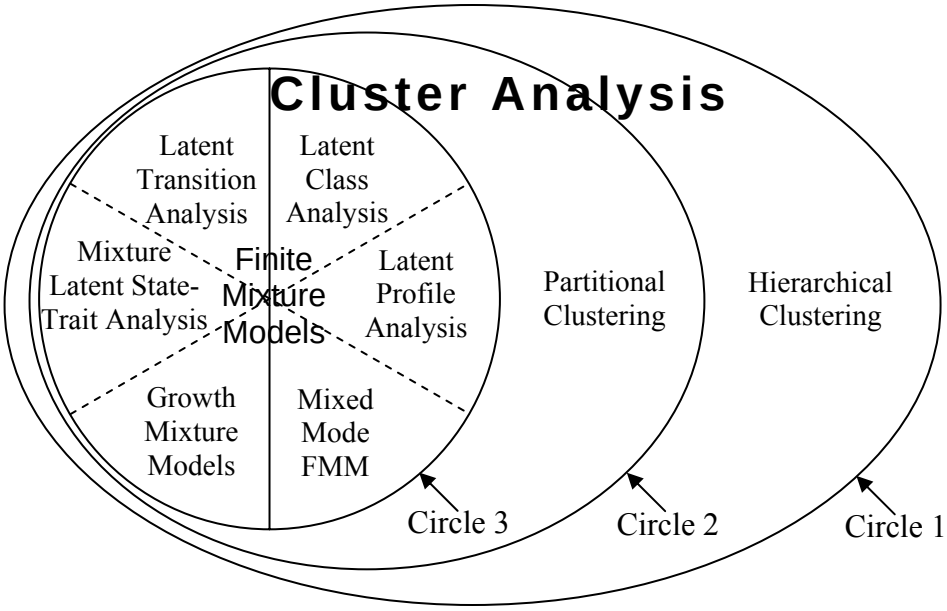

Figure 1. Diagram illustrating relationships between models.

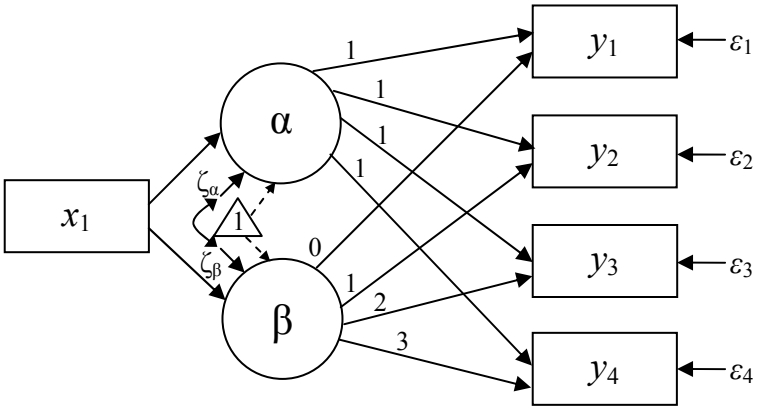

Figure 2. Linear latent growth curve model with four repeated measures and one covariate.

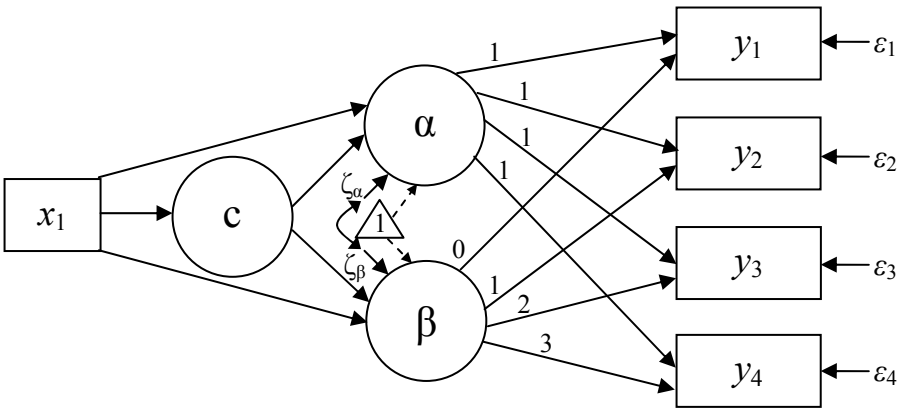

Figure 3. Linear growth mixture model with four repeated measures and one covariate predicting both growth factors and latent class.

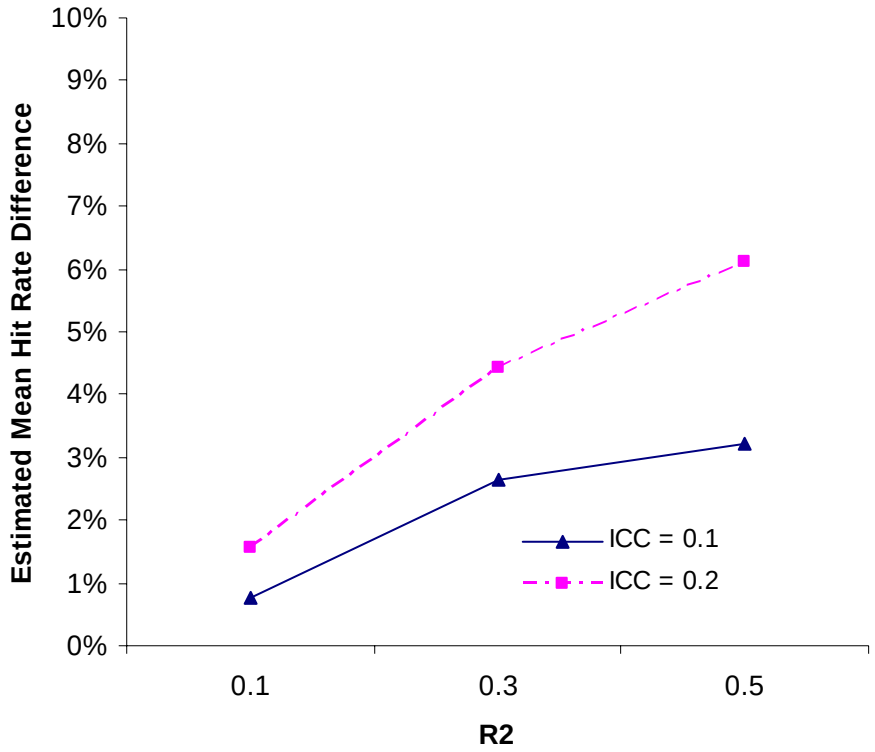

Figure 4. Interaction effect between ICC and R^2 for estimated mean hit rate difference between true and false models.

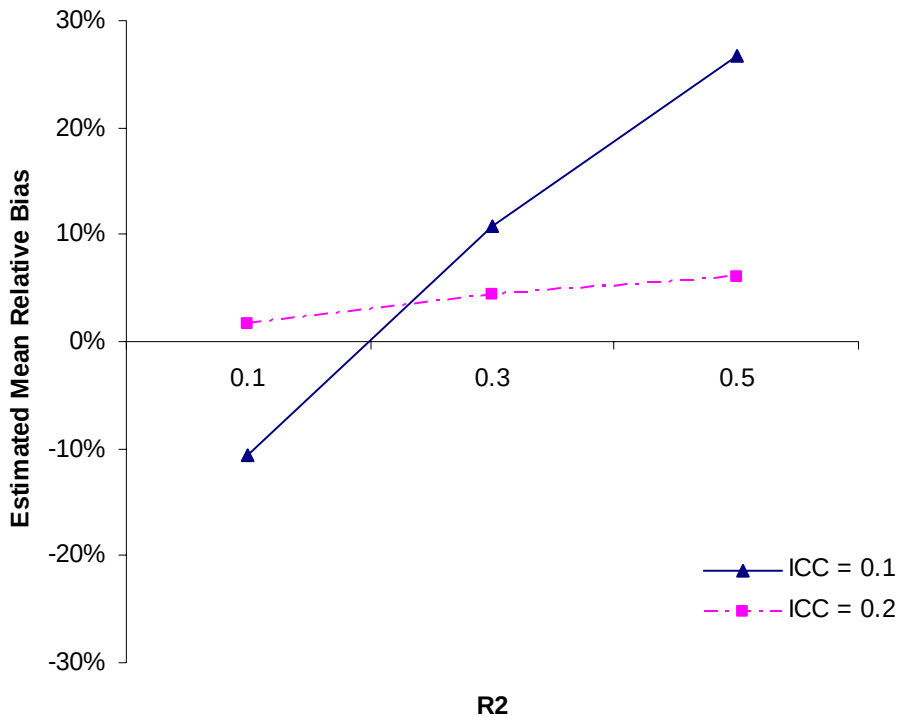

Figure 5. Interaction effect between ICC and R^2 for estimated mean relative bias for level-1 variance estimate under false model.

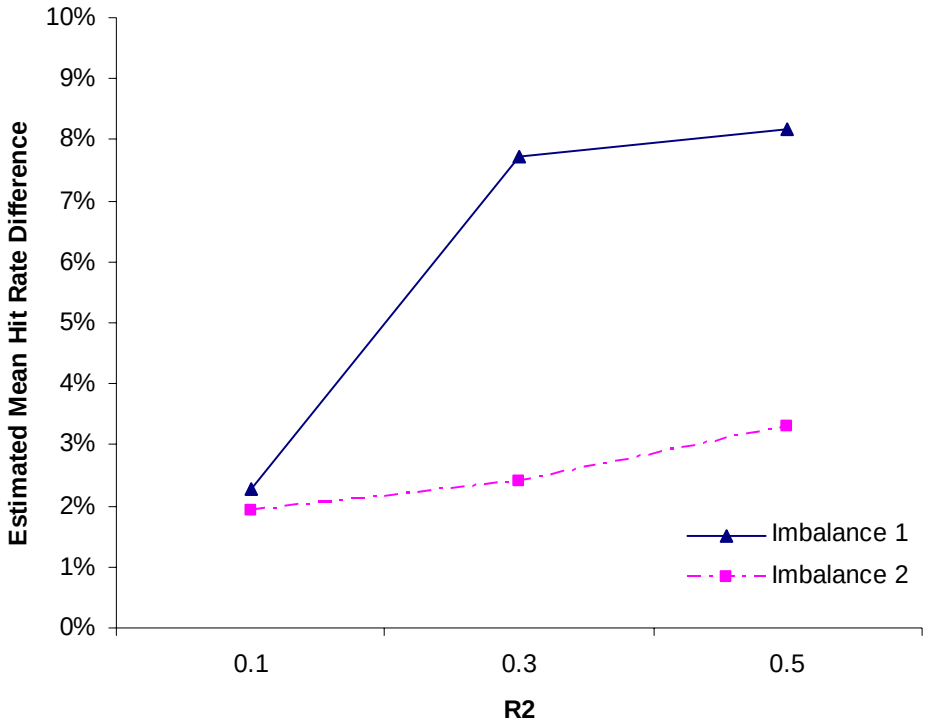

Figure 6. Interaction effect between imbalance type and R^2 for estimated mean hit rate difference under true and false models.

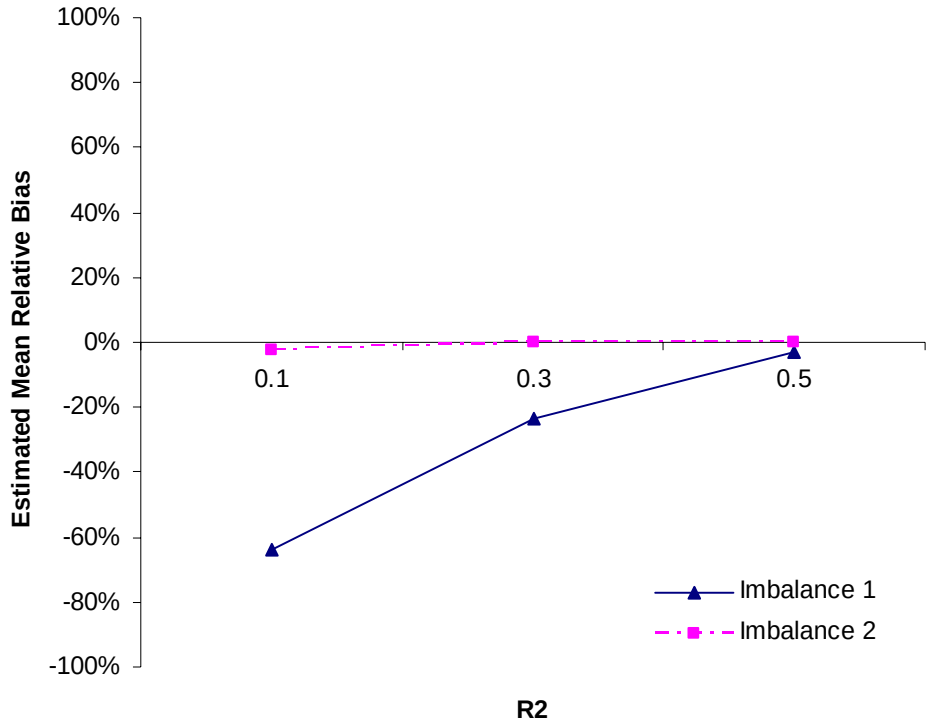

Figure 7. Interaction effect between imbalance type and R^2 for estimated mean relative bias for Class 1 under true model.

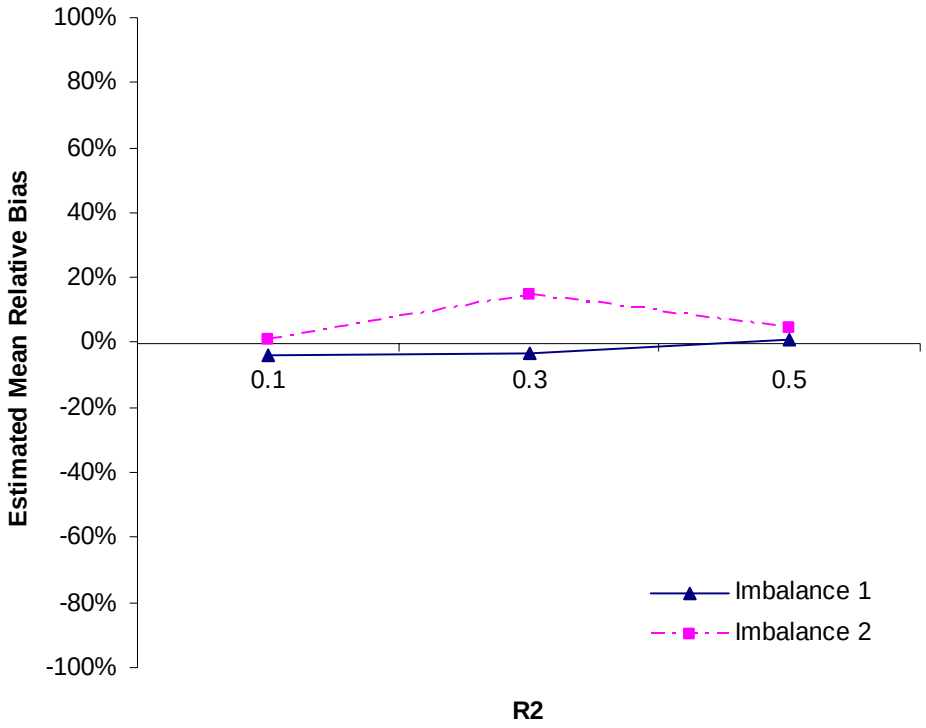

Figure 8. Interaction effect between imbalance type and R^2 for estimated mean relative bias for Class 2 under true model.

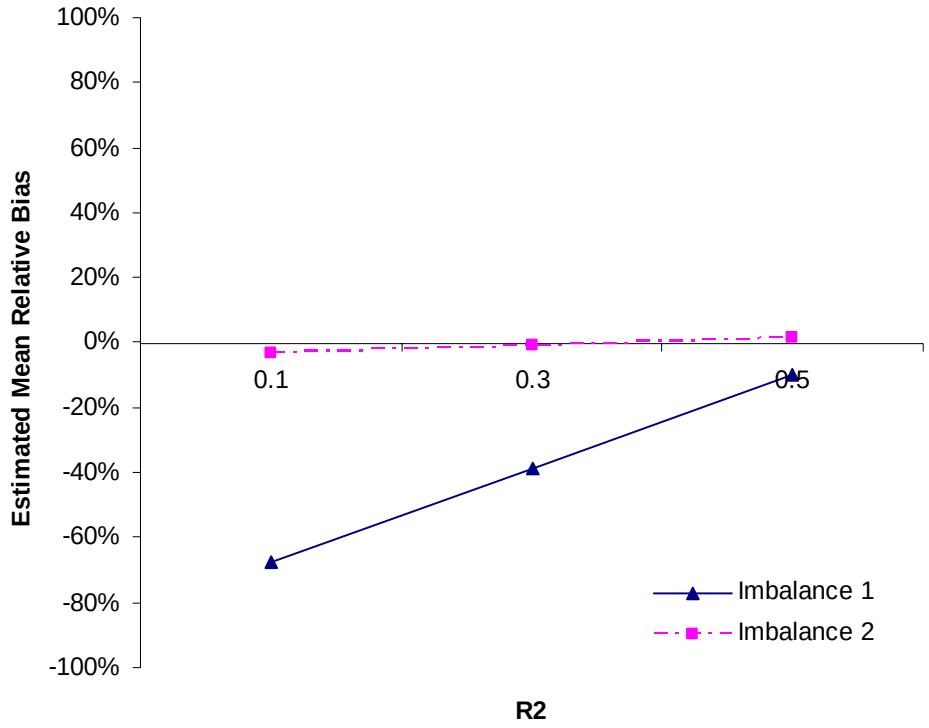

Figure 9. Interaction effect between imbalance type and R^2 for estimated mean relative bias for Class 1 under false model.

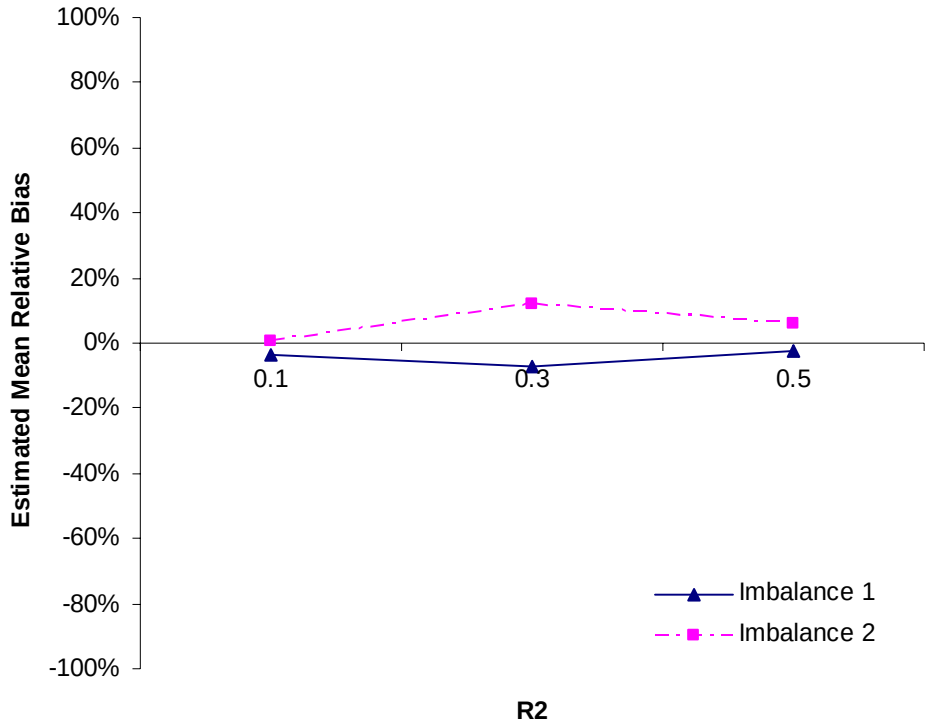

Figure 10. Interaction effect between imbalance type and R^2 for estimated mean relative bias for Class 2 under false model.

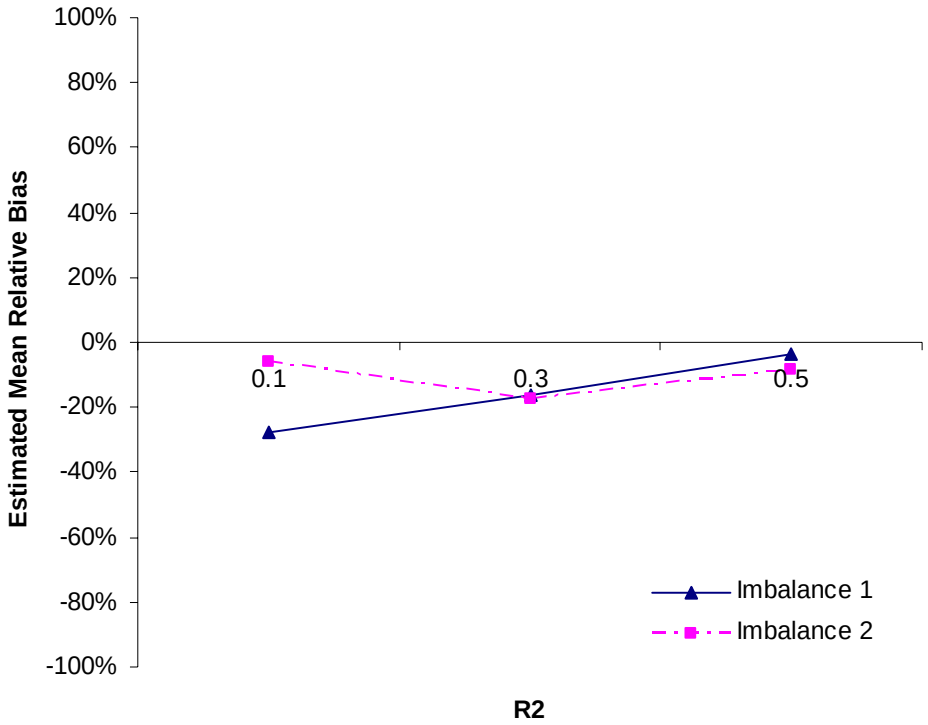

Figure 11. Interaction effect between imbalance type and R^2 for estimated mean relative bias for Class 1 level-1 variance under true model.

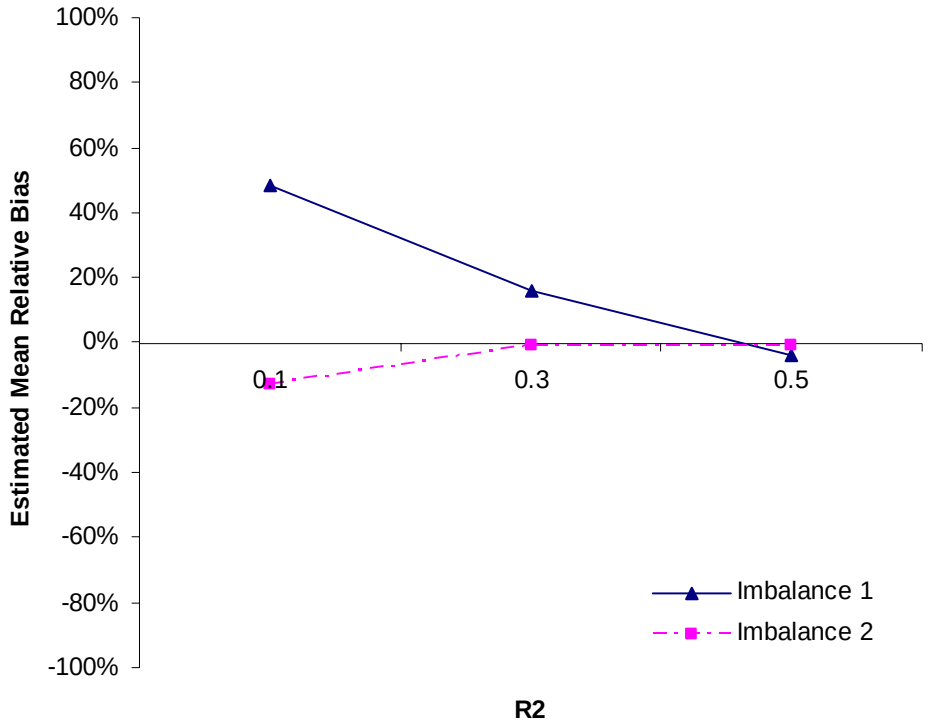

Figure 12. Interaction effect between imbalance type and R^2 for estimated mean relative bias for Class 2 level-1 variance under true model.

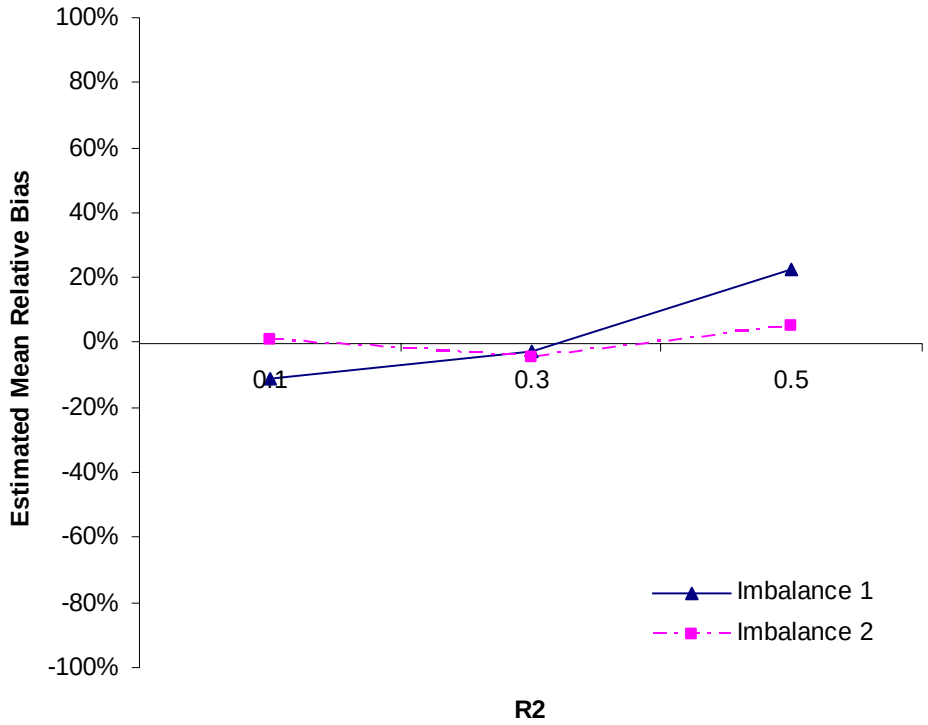

Figure 13. Interaction effect between imbalance type and R^2 for estimated mean relative bias for Class 1 level-1 variance under false model.

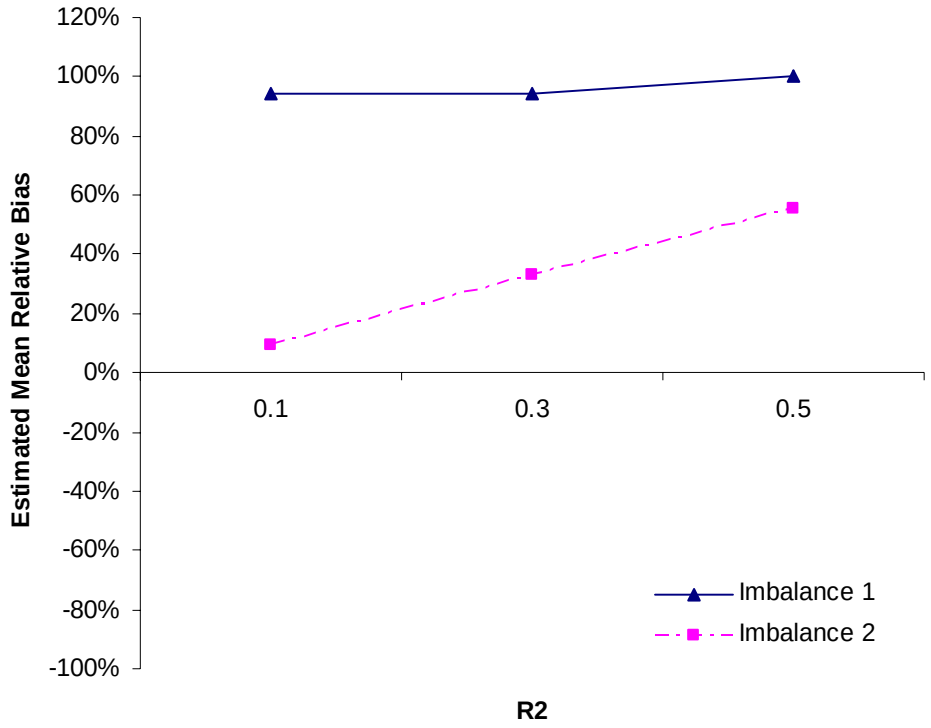

Figure 14. Interaction effect between imbalance type and R^2 for estimated mean relative bias for Class 2 level-1 variance under false model.

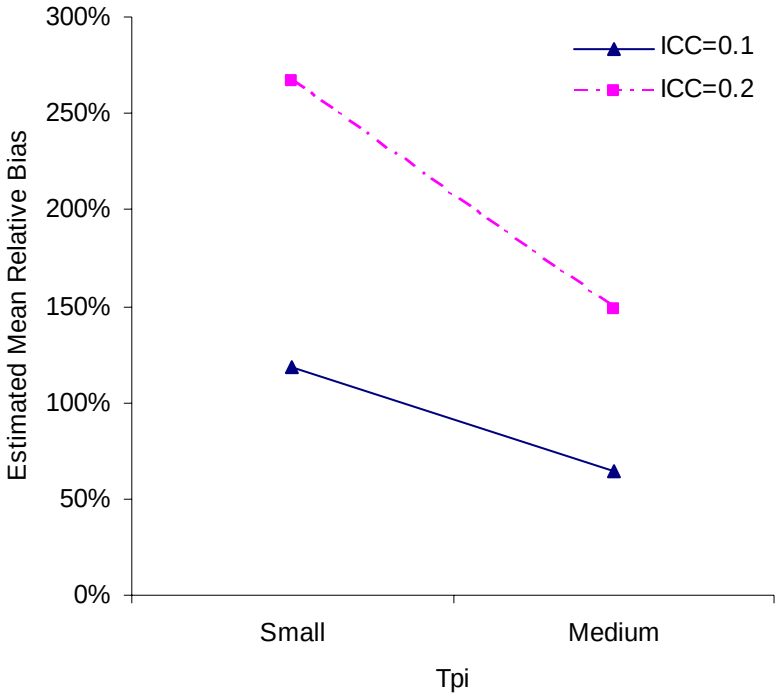

Figure 15. Interaction effect between ICC and τ_π for estimated mean relative bias of intercept variance under false model.

VITA

Name:	Qi Chen
Address:	Department of Educational Psychology Texas A&M University College Station, TX 77843-4225
Email Address:	jochenqi@gmail.com
Education:	B.A., English Language and Literature, Nanjing University, 1999 M.A., English: Applied Linguistics, Nanjing University, 2002